The New Genetic Based Method with Optimum Number of Super Node for Fault Tolerant System in Heterogeneous Wireless Sensor Network

Nima Jafari Navimipour, *Member, IAUT,* and Amir Masoud Rahmani

Abstract—In this paper the new method for determining the number of super node in heterogeneous wireless sensor network based on evolutionary algorithms are presented. The network consisting of several resource-rich supernodes, used for data relaying, and a large number of energy-constrained wireless sensor nodes. The main contribution of this paper is to reach an optimum trade off between number of super node and efficiency. Simulation results show that our algorithm can quickly find a good solution.

Index Terms—Heterogeneous Wireless Sensor Network, Super Node, Genetic Algorithm, Fault Tolerant System

I. INTRODUCTION

Advances in wireless sensor networks make many of the impossible possible. Roadway safety warning [1], habitat monitoring [2], smart classroom [3], etc., are prosperous applications tied to our daily life. Such networks rely on the collaboration of thousands of resource-constrained error-prone sensors for monitoring and control. One of the most contemporary challenges is to design efficient methods for exploiting the new technology of wireless sensor networks (WSN). A WSN consists of a large number of sensor nodes deployed over a certain area, providing real-time data about certain phenomena [4, 5, and 6]. The deployment of a WSN can be random (for example, dropping sensors in a hostile terrain or a disaster area) or deterministic (for example, placing sensors along a pipeline to monitor pressure and/or temperature, and boundary surveillance). WSN divided to two important types: homogenous wireless sensor network and heterogeneous wireless sensor network. In homogenous WSNs all nodes in the network have the same power, resources, quality and etc, but heterogeneous WSNs consisting of two types of wireless devices: resource-constrained wireless sensor nodes deployed randomly in a large number and a much smaller number of resource-rich super nodes placed at known locations [7]. The super nodes have two transceivers: one connects to the WSN, and the other connects to the super node network. The upper node network provides better Quality of Service (QoS) and is used to quickly forward sensor data packets to the user. A study by Intel [8] shows that using a heterogeneous

architecture results in improved network performance such as a lower data-gathering delay and a longer network lifetime. Hardware components of the heterogeneous WSNs are now commercially available [9].

The super nodes are more expensive than sensor nodes, so the minimizing the number of super nodes is very important. The trade off between minimizing the number of super node and the efficiency must be considered. The efficiency of this network defined as reducing the total power and the communication distance between nodes. The main contribution of this paper is to introduce the new genetic base approach for solving this problem. The traditional methods for solving this NP-Hard problem are very slow and inefficient so we need the indeterminist methods. The Genetic Algorithm (GA) is a randomly method for solving the problems that widely use in many areas such as Grid Scheduling [10], Telecom Services [11], Timetabling Problems [12] and etc. GA inspired from natural evolutionary process. However, this way hasn't guaranteed the optimum response but always find the solution that is close to optimum or optimum. In the next section the brief survey on related works explained. Section 3 discusses the Genetic Algorithm. In section 4 the network model presented. Section 5 presents the new method and the last section is conclusion.

II. RELATED WORK

The benefits of using heterogeneous WSNs, containing devices with different capabilities, have been presented recently in the literature. In [13], the authors reported that when the nodes properly deployed, heterogeneity can triple the average delivery rate and provide a five-fold increase in the network lifetime. The work in [14] introduces another type of heterogeneous WSNs called actor networks, consisting of sensor nodes and actor nodes. The role of actor nodes is to collect sensor data and perform appropriate actions. This paper presents an event-based coordination framework using linear programming and a distributed solution with an adaptive mechanism to trade off energy consumption for delay when event data has to be delivered within a specific latency bound.

This work is motivated by Heinzelman et al's paper [15] "Energy-Efficient Communication Protocol for Wireless Micro-sensor Networks" which describes a new protocol called LEACH. They compare the performance of LEACH with direct communication and MTE. They use a pre-determined optimal number of clusters (5% of the total number of nodes) in their simulations. Tillett et al [16]

Nima Jafari Navimipour is member of Young Researchers Club, Tabriz Branch, Islamic Azad University, Tabriz Branch, Tabriz, Iran, phone: +989144021694; fax: +984115552322; e-mail: Jafari@iaut.ac.ir and Jafari_n@ymail.com

propose the PSO (Particle Swarm Optimization) approach to divide a sensor node field into groups of equal sized groups of nodes. PSO is an evolutionary programming technique that mimics the interaction of ants or termites to find a good solution. Although partitioning into equal sized clusters balances the energy consumption of cluster heads, this method is not applicable to some networks where nodes are not evenly distributed.

Ostrosky et al [17] address a somewhat different partitioning problem: Given n points in a large data set, partition this data set into k disjoint clusters so as to minimize the total distance between all points and the cluster-heads to which they belong. The authors use a polynomial-time approximation scheme to solve the problem.

In [7] three methods have been introduced for fault tolerant topology in heterogeneous WSNs. The authors explain the network that has several sensor nodes and super nodes. In this paper the number of super node is fixed. We would like to solve this problem where we do not know the number of super node in advance. Our approach uses a GA to determine both the number and location of the super nodes that minimizes the communication distance and total consumption power in a sensor network. The next section briefly presents the genetic algorithm.

III. GENETIC ALGORITHM

GA is an efficient searching tool that was invented by John Holland [18].The genetic algorithm has great application for optimization of complicated problems particularly in where is not adequate information about search space. Although, considering that, genetic algorithm is not guarantee best possible solution, but normally, would provide optimum or partly optimum solution by suitable approximate at short time. For solving any problem by genetic algorithm, seven components must be defined [19].

- **Representation (definition of individual):** represents each chromosome in the real world. A chromosome is a set of parameters which define a proposed solution to the problem that the genetic algorithm is trying to solve.
- **Fitness function:** These function shows the fitness of each chromosome. It is used to evaluate the chromosome and also controls the genetic operators.
- **Population:** The role of the population is to hold possible solution.
- **Parent selection mechanism:** The role of parent selection is to distinguish among individuals based on their quality, in particular, to allow the better individuals to become parents of the next generation.
- **Reproduction:** The reproduction operator is based on the Darwinian notion of "survival of the fittest". Individuals taking part in successive generations are obtained through a reproduction process or evolution operation. Individual strings are copied into a mating pool according to their respective fitness values. The higher the fitness values of the strings, the higher the probability of contributing one or more offspring in the next generation.
- **Crossover operators:** Recombination operator selects two or more chromosomes and then produces two new children from them. It aims at mixing up genetic information coming from different chromosomes to make a new individual.
- **Mutation operators:** Mutation operator selects one chromosome and then produces one new child from it by a slight change over the parent.
- **Survivor selection mechanism:** The role of survivor selection is to distinguish among individuals based on their quality. This mechanism survives the individual among the passing from one generation to the next generation.
- **Termination Condition:** The condition to ending the running of genetic algorithm.

After defining these components, some probability components must be defined as follow:

- **Probability of Mutation (P_m):** P_m is the probability that one chromosome selected for mutation. This probability acts on chromosomes.
- **Probability of Crossover (P_c):** P_c is the probability that one chromosome selected for crossover. This probability acts on chromosomes.
- **Probability of Gene flipping (P_{gf}):** P_{gf} is the probability that which gene on the selected chromosome (selecting the chromosome with P_m) mutated. This probability acts on genes.
- **Population Size:** population size is the number of solutions that hold in on generation, this number almost is constant.

The next section is about network model of the problem.

IV. NETWORK MODEL

We consider a heterogeneous WSN consisting of M super nodes and N sensor nodes, with M<< N. We are interested in sensor-super node communications only. We represent the network topology with an undirected weighted graph $G=(V, E, c)$ in the 2D plane, where $V=\{n_1,n_2,...,n_N,n_{N+1},...,n_{N+M}\}$ is the set of nodes, and E is the set of edges. The first N nodes in V are the sensor nodes, and the last M nodes are the super nodes.

We define the set of edges $E=\{(n_i,n_j)/dist(n_i,n_j)\leq R_{max}\}$ where $dist()$ is the distance function.

The cost function $c(u,v)$ represents the power requirement for both nodes u and v to establish a bidirectional communication link between u and v.

We assume that each node has a unique id such as the MAC address and that each node is able to gather its own location information by using one of the localization techniques for wireless networks such as [20].

The reachable neighborhood is the set of nodes that node n_i can reach by using the maximum transmission range R_{max}.

Figure 1 is an example of direct transmission where each sensor transmits messages directly to the sink.
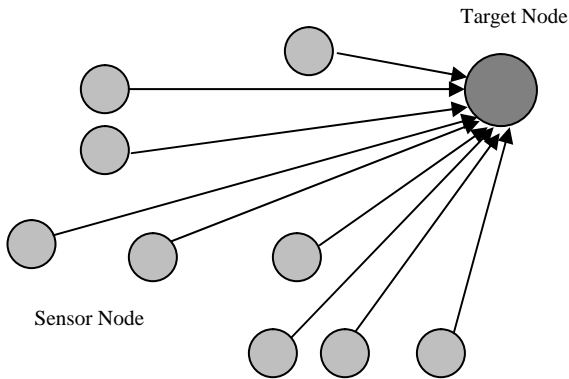
Figure.1   An example of direct transmission

Direct transmission networks are very simple to design but can be very power-consuming due to the long distances from sensors to the target node. Alternative designs that shorten or minimize the communication distances can extend network lifetimes.

The use of super nodes for transmitting data to a target node leverages the advantages of small transmit distances for most nodes, requiring only a few nodes to transmit far distances to the target node[15]. The super nodes gather the data and send it directly to the target node. This model can greatly reduce communication costs of most nodes because they only need to send data to the nearest Super node, rather than directly to a target node that may be further away.

In this paper, we assume the sensor network is static. Sensors are deployed in a remote inhospitable environment and are far away from the target node which is usually positioned in a safe place. All nodes are assumed to have the capabilities of a super node and the ability to adjust their transmission power based on transmission distance. Each sensor's position can be precisely measured by GPS 1 devices.

Minimizing the total distance in the WSN is an NP-hard [21] problem. For a given network topology, it is difficult to find the optimal number of super node and their locations. Consider a 80-node example, to perform an exhausted search of all possible solutions requires $C^1_{80}+C^2_{80}+....+C^{80}_{80}=2^{80}-1$ different combination which is far too large to be handled by existing computer resources.

A genetic algorithm is an efficient search algorithm that mimics the adaptive evolution process of natural systems. It has been successfully applied to many NP-hard problems such as multi-processor task scheduling, optimization, and traveling salesman problems. We propose a GA to the problem of minimizing the total communication distance in a sensor network to efficiently reduce energy consumption and maximize the lifetime of the network. Figure 2 shows an example of network model.
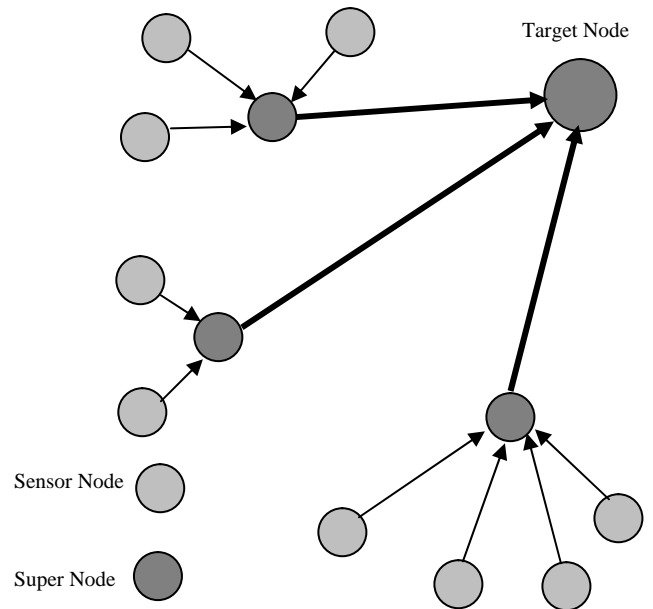

Figure.2   an example of data gathering in heterogeneous WSN

V.   NEW METHOD

We would like to use a GA to optimize the number of super nodes and sensor connections for an arbitrary network. Once super nodes are selected, each regular node connects to its nearest super nodes. Each node in a network is either super nodes or a sensor node that connected to the nearest super nodes. Each super node collects data from all sensors within its area and each super node directly sends the collected data to the target node. In the next ten subsections the GA are explained for heterogonous WSNs.

A.   Representation

In this method the binary representation is used. Each bit corresponds to one sensor node or super node. A "1"means that corresponding sensor is a super node; otherwise, it is a sensor node. In the figure 3 one example of the chromosome illustrated.

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
1	0	0	1	0	0	1	0

Figure.3   one chromosome with three super nodes

Nodes S1, S4 and S7 are super nodes and S2, S3, S5, S6 and S8 are sensor nodes.

B.   Fitness Function

The total transmission distance is the main factor we need to minimize. In addition, the number of super nodes can factor into the function. Given the same distance, fewer super nodes result in greater energy efficiency as super nodes drain more power than sensor nodes. Thus, each individual is evaluated by the following combined fitness components:

$Fitness = (w \times (D\text{-}I) + (1\text{-}w) \times (N - S_i)) / 100$

$I = I_n + I_s$

Where D is the total distance of all nodes to the target node; I_n is the distances from sensor nodes to super node. I_s is the

sum of the distances from all super nodes to the target node; S_i is the number of super nodes; N is the total number of nodes; and w is a pre-defined weight. Except for I and Si, all other parameters are fixed values in a given topology. The shorter the distance, or the lower the number of super nodes, the higher the fitness value of an individual is. Our GA tries to maximize the fitness value to find a good solution.

The value of w ($0 \leq w \leq 1$) is application-dependent. It indicates which factor is more important to be considered: distance or the cost incurred by super nodes. At one extreme, if w=1, we optimize the network only based on the communication distance. If w=0, only the number of super nodes is considered.

C. Population

The initial population consists of randomly generated individuals. The population size is varied and determined based on the number of nodes in the network. In this method the population size considered equal to the number of nodes.

D. Parent Selection Mechanism

This method for selecting the parent, evaluated the fitness of each chromosome. Then selects 10 chromosomes randomly, then for mutation operating this method selects the 5 best chromosomes, then selects one chromosome randomly. For crossover operating, this method select one chromosome from 5 best chromosomes randomly and another chromosome selecting from 9 remaining chromosome randomly.

E. Reproduction

There are many ways to implement a satisfactory reproduction operator. One commonly adopted implementation called fitness proportionate selection assumes that each individual has an associated fitness value. The GA generates new populations by reproducing (making three copies of) chromosome under weighted probabilities in proportion to their fitness. Thus, a higher fit chromosome is more likely to survive from generation to generation.

F. Crossover

This method uses the one point crossover for recombination operator.

The crossover operation takes place between two consecutive chromosomes with pc = 75%. These two individuals exchange portions that are separated by the crossover point. For example suppose that two chromosome that specified in figure 4 selected for crossover. The crossover point is after the fourth gene.

| First Parent | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| Second Parent | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |

Figure.4 two chromosomes for crossover operation.

In this step two chromosomes exchange the genes base on the crossover point and two children produced. Figure 5 illustrates these children.

| First Child | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Second Child | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |

Figure.5 two children that produced after the crossover operation.

After crossover, if a sensor node becomes a super node; all other sensor nodes should check if they are nearer to this new

super node. If so, they switch their membership to this new super node. If a super node becomes a sensor node, all of the nodes that communicate with this super node must find new super node. Every node is either a super node or a sensor node in the network.

G. Mutation

The mutation operator is applied to each bit of a chromosome with pm = 2% and pgf = 1/n where n is the length of chromosome or number of the nodes in the network. The bitwise mutation is used here. The pgf when applied, a bit whose value is 0 is mutated into 1 and vice versa. An example of mutation is in figure 6. In this figure the chromosome that illustrated in figure 3 mutated.

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
1	0	0	**0**	0	0	1	**1**

Figure.6 the mutated chromosome

As shown in figure 6 the genes that placed in the 4th and 8th place, flipped.

After mutation, if a sensor node becomes a super node; all other sensor nodes should check if they are nearer to this new super node. If so, they switch their membership to this new super node. If a super node becomes a sensor node, all of the nodes that communicate with this super node must find new super node. Every node is either a super node or a sensor node in the network.

H. Survivor Selection Mechanism

After that the middle population size (mating pool size) is the 75% of the population size, the 25% of the best chromosome of the old population directly copped to the new population. The remaining chromosome, 75% of population size, replaced with the new chromosome from the middle population.

I. Termination Condition

In this method, our algorithm running until no improvement in the fitness of the best member of the population has been observed for 20 generation.

J. Pseudo Code of Method

1. [Initialize]
2. Repeat steps 3 to 7 until algorithm is convergent
3. Initial Population
4. Fitness Function
5. Perform Crossover
6. Perform Reproduction
7. Perform Mutation
8. Show the Best Chromosome

K. GA Summary

In the previous subsections, we described the new method's components. Putting this all together, we obtain as GA as summarized in Table 2.

TABLE.1 Summary of genetic algorithm components

Representation	Binary representation
Recombination	"one point" crossover
p_c	75%

Mutation	bitwise mutation
p_m	3%
p_{gf}	1/n
Parent selection for crossover	First, selects 10 chromosomes randomly, and then selects 2 chromosomes from 5 best chromosomes randomly.
Parent selection for mutation	First, selects 10 chromosomes randomly, and then selects one chromosome from 5 best chromosomes randomly.
Survival selection	Replace worse
Population size	n
Initialization	Random
Termination condition	No improvement in Last 20 generations.

VI. RESULT AND SIMULATION

In this section the simulation result for this method presented. MATLAB used for this simulation. The number of sensor is *50*. The area that the sensor node deployed is *20×20*. Table1 show the simulation result when *w = 1* and the target sensor is located at the upper left corner.

In table1 the first column indicates the sensor number, X and Y show the location of sensor in the network, fourth column show the super node number and the last column is type of sensor.

This table show that *6* super nodes needed for *50* nodes in *20 × 20* area. The other nodes by considering the minimum power connected to one of the super nodes.

TABLE.2 the simulation result when *w = 1* and the target sensor is located at the upper left corner.

Node #	X	Y	Super Node	Type
1	0	0	-------	Target Node
2	7	6	35	Sensor Node
3	6	16	3	Super Node
4	19	16	20	Sensor Node
5	10	8	50	Sensor Node
6	18	2	6	Super Node
7	10	10	50	Sensor Node
8	4	15	3	Sensor Node
9	4	19	29	Sensor Node
10	20	4	6	Sensor Node
11	5	17	29	Sensor Node
12	15	15	20	Sensor Node
13	13	9	50	Sensor Node
14	19	1	6	Sensor Node
15	5	7	29	Sensor Node
16	5	19	3	Sensor Node
17	4	20	29	Sensor Node
18	10	16	3	Sensor Node
19	5	2	35	Sensor Node
20	18	17	20	Super Node
21	13	10	50	Sensor Node
22	1	18	29	Sensor Node
23	16	4	6	Sensor Node
24	16	16	20	Sensor Node
25	16	18	20	Sensor Node
26	9	20	3	Sensor Node
27	6	16	3	Sensor Node
28	4	6	35	Sensor Node
29	3	18	29	Super Node
30	12	10	50	Sensor Node
31	17	17	20	Sensor Node
32	0	18	29	Sensor Node
33	18	15	20	Sensor Node
34	10	8	50	Sensor Node
35	5	4	35	Super Node
36	17	18	20	Sensor Node
37	19	16	20	Sensor Node
38	5	17	29	Sensor Node
39	7	8	35	Sensor Node
40	7	3	35	Sensor Node
41	8	18	3	Sensor Node
42	11	14	50	Sensor Node
43	5	17	29	Sensor Node
44	15	8	50	Sensor Node
45	15	4	6	Sensor Node
46	19	15	20	Sensor Node
47	4	5	35	Sensor Node
48	2	4	35	Sensor Node
49	5	18	29	Sensor Node
50	11	9	50	Super Node

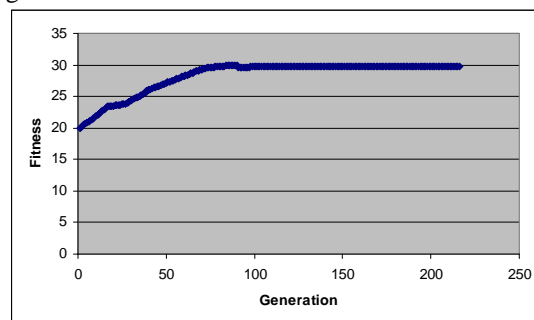Figure 7 illustrates the fitness value during of the generation.

Figure.7 the fitness value during the generation

As shown in figure 7, fitness value increased up to the 80 generation and then almost is constant.

Figure 8 illustrates the simulation result for decreasing the total distance during the generation.
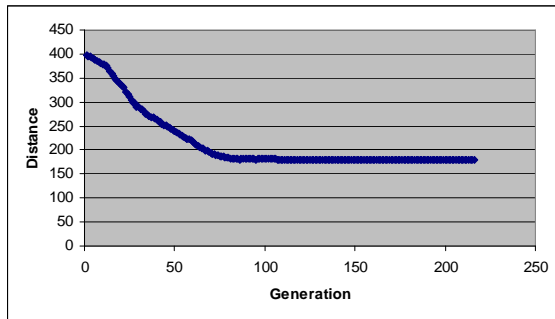
Figure.8   decreasing the total distance during the generation

As shown in this figure, the total distance minimized in first 80 generation.

These results show that our method find the best solution very fast.

VII. CONCLUSION

In this paper the new method for defining the number of super node in heterogeneous wireless sensor network are presented. Genetic algorithm as a member of evolutionary algorithms is used for solving this problem. This algorithm begins by randomly selecting nodes in a network to be super node. This algorithm is able to find an appropriate number of super nodes and their locations. The simulation results show that this method can find the best solution very fast.

REFERENCES

[1] K. Xing, X. Cheng, and M. Ding, Safety Warning Based on Roadway Sensor Networks, submit to IEEE Wireless Communications and Networking Conference 2005.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, Wireless Sensor Networks for Habitat Monitoring, ACM WSNA'02, Atlanta GA, September 2002.

[3] S. S. Yau, S. K. S. Gupta, F. Karim, S. I. Ahamed, Y. Wang, and B. Wang, Smart Classroom: Enhancing Collaborative Learning Using Pervasive Computing Technology, Proc. of 6th WFEO World Congress on Engineering Education and Second ASEE International Colloquium on Engineering Education (ASEE), June 2003, Nashville, Tennessee.

[4] J. Agre and L. Clare, "An Integrated Architecture for Cooperative Sensing Networks," IEEE Computer, vol. 5, pp, 106-108, 2000.

[5] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: a Survey," Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.

[6] David Culler, D. Estrin and M. Srivastava, "Overview of Sensor Networks," IEEE Computer, August, pp. 41-49, 2004.

[7] Mihaela Cardei, Shuhui Yang, "Algorithms for Fault-Tolerant Topology in Heterogeneous Wireless Sensor Networks," IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 19, NO. 4, APRIL 2008

[8] Heterogeneous Networks with Intel XScale, http://www.intel.com/ research/ exploratory/ heterogeneous.htm, 2007.

[9] Crossbow Mica2 Motes and Stargate-Xscale, http://www.xbow.com, 2007.

[10] Nima Jafari Navimipour, Leili Mohammad Khanli, "The LGR Method for Task Scheduling in Computational Grid,"ICACTE, IEEE, pp.1062-1066, 2008 International Conference on Advanced Computer Theory and Engineering, Phuket, Thailand, 2008

[11] Linda Murphy, Hoda S. Abdel-Aty-Zohdy, M. Hashem-Sherif, "A Genetic Algorithm Tracking Model for Product Deployment in Telecom Services," pp 1729-1732, IEEE, 2005

[12] MILENA KAROVA, "Solving Timetabling Problems Using Genetic Algorithms," 27th Int'l Spring Seminar on Electronics Technology, pp 96-98, IEEE, 2004

[13] M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, and S. Singh, "Exploiting Heterogeneity in Sensor Networks," Proc. IEEE INFOCOM, 2005.

[14] T. Melodia, D. Pompili, V.C. Gungor, and I. Akyildiz, "A Distributed Coordination Framework for Wireless Sensor and Actor Networks," Proc. ACM MobiHoc, 2005.

[15] [15] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-Efficient Communication Protocol for Wireless Micro-sensor Networks. In Proceedings of the Hawaii International Conference on System Science, Maui, Hawaii,2000

[16] J. Tillett, R. Rao, F. Sahin, and T.M. Rao. Cluster-head Identification in Ad hoc Sensor Networks Using Particle Swarm Optimization. In Proceedings of the IEEE International Conference on Personal Wireless Communication, 2002.

[17] R. Ostrosky and Y. Rabani. Polynomial-Time Approxiamation Schemes for Geometric Min-Sum Median Clustering. Journal of the ACM, Vol. 49, No. 2, 2002, pp 139-156.

[18] Holland, J. H. (1975) "Adaption in Natural and Artificial Systems". The University of Michigan Press, Ann Arbor, MI, 1975

[19] A.E.Eiben, J.E.Smith, "Introduction to Evolutionary Computing", Springer, 2004

[20] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "TPS: A Time-Based Positioning Scheme for Outdoor Wireless Sensor Networks," Proc. IEEE INFOCOM, 2004.

[21] P. Pardalos and H. Wolkowicz. Quadratic Assignments and Related Problems, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 16,1994.

Nima Jafari Navimipour received his B.S. in computer engineering, software engineering, from Islamic Azad University, Tabriz Branch, Tabriz, Iran, in 2007, the M.S. in computer engineering, computer architecture, from Islamic Azad University, Tabriz Branch, Tabriz, Iran, in 2009. From 2007, he worked as a researcher with the Young Researchers Club, Tabriz Branch, Islamic Azad University, Tabriz Branch.

He is the author/co-author of more than 15 publications in technical journals and conferences. His research interests include Traffic Control, Traffic Modeling, Computational Intelligence, Evolutionary Computing, Computational Grid, and Wireless Networks.

Amir Masoud Rahmani received his B.S. in computer engineering from Amir Kabir University, Tehran, in 1996, the M.S. in computer engineering from Sharif University of technology, Tehran, in 1998 and the PhD degree in computer engineering from IAU University, Tehran, in 2005. He is an assistant professor in the Department of Computer and Mechatronics Engineering at the IAU University.

He is the author/co-author of more than 60 publications in technical journals and conferences. He served on the program committees of several national and international conferences. His research interests are in the areas of distributed systems, ad hoc and sensor wireless networks, scheduling algorithms and evolutionary computing.