

An Energy-aware Iterative Sampling Framework for Data Gathering in Wireless Sensor Networks

Jun Wang^{1,2}, Zhenglu Wang¹, Yong Cheng² and Yongsheng Zhu²

¹Dept. of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044, China

²Network Information Center, Nanjing University of Information Science & Technology, Nanjing 210044, China

wangjun@nuist.edu.cn, wzl270mail@gmail.com

Abstract

Large numbers of nodes are often densely deployed to deliver the desired environmental attributes to the sink in Wireless Sensor Networks (WSNs), so there is a high spatial correlation among the readings of close sensor nodes. Given a certain requirement for accuracy, only part of the sensor nodes should be required to transport the data to sink. We proposed an Energy-aware Iterative Sampling Framework (EISF) for data gathering to reduce the total number of transmissions by exploiting the correlation. In our method, all nodes in a WSNs compete for reporting nodes with energy-related probability and each nonreporting node autonomously determines whether its own readings are redundant or not by utilizing the overheard packets transmitted by the nearby reporting nodes for each epoch. The redundant nodes will be put into sleep mode. After a limited number of iterations, our algorithm can select a set of sampling nodes to transport data with accuracy guarantees. The results of simulation experiments using the real data demonstrate that our proposed approach is effective in prolonging the network life.

Keywords: energy-aware, data gathering, iterative sampling, spatial correlation, wireless sensor networks

1. Introduction

Recent advances in wireless communications and electronics have enabled WSNs to be applied in a wide range of applications, such as battlefield surveillance, environmental monitoring, and disaster relief. A large number of sensor nodes are often deployed to collect environmental surveillance data, temperature, humidity, and *etc.* However, spatial correlation usually exists among the readings of close nodes because of redundant deployment. The measures of a node may be predicted from the readings of its nearby nodes with high confidence. Therefore, the “over-sampling” problem will occur when the nodes transporting readings to sink in one cycle is enough. Sensor nodes are battery powered and sending the large mount of data rapidly depletes their energy. To achieve energy efficiency, we hope to appoint a part of nodes as sampling nodes to transmit data by reducing the spatial correlation and the other node’s data can be interpolated at the sink with the known data.

In this paper, we use two strategies to improve the efficiency of the network. The first one is to use a few nodes to transport data and enforce error control; The second is to increase the probability of being sampling nodes for the nodes with high energy. We propose an Energy-aware Iterative Sampling Framework for data gathering (EISF) to

select a subset of sampling nodes iteratively through an energy-aware competitive mechanism. In data gathering phase, only the sampling nodes forward the data to sink. Firstly, we assume all nodes know its location and the sink's and transport data to sink periodically. In every cycle, some nodes will be selected as sampling nodes and then they transmit the readings to their neighbors. The other nodes those failed to be selected will determine whether its reading is redundant or not based on the information transmitted by sampling nodes. If a node's redundancy is no more than the error bound set in advance, it transmits the confirming message to the transmitters and then goes to sleep until next cycle. The redundancy of readings among nodes is estimated by Inverse Distance Weighting interpolation (IDW). All sampling nodes transmit their readings and the confirming message to sink after a limited number of iterations. Nodes will calculate the probability according to their residual energy in each iteration. The probability decides whether a node will be a sampling node or not. So the node with more energy will be more likely to be a sampling node. The energy consumption of the whole network will be more effective. The sink will use the same interpolating method to interpolate the missing data using the sampling nodes' readings.

Our sampling framework for data gathering only studied the spatial correlation among nodes. The temporal correlation of monitoring data was exploited in [1, 3, and 7] and can be applied into our method. The rest of the paper was organized as follows: In Section 2, we described the existing methods for data gathering and discussed about their advantages and disadvantages. In Section 3, we described our framework and the algorithms in more detail. In Section 4, we simulated our method and compared the performances of it with other two methods' using real experimental data. In the last section, we concluded the paper and described some further works.

2. Related work

A very important problem in applications of WSNs is how to save the energy of nodes and prolong the life of the network. Sending a message is a costly energy consuming operation. Energy-efficient frameworks for data aggregation can decrease the energy consumption on communication.

Researchers have studied some kinds of frameworks for data gathering for WSNs [6-18]. The methods in [9, 11, and 17] are model-driven sampling frameworks, using an approach similar to data prediction. In general, the sink and nodes have the same prediction model. The models are used to decide the data of a node is redundant or not, and sample the nodes to transmit data. However, the models need to be trained and preserved frequently that consumes a lot of power. In the study [17], the researchers use a dynamic probabilistic model to minimize communication by comparing the difference between prediction and measurement. The sink keeps the prediction instead of real reading of a node when the deviation is small enough. The centralized methods in [3, 8] waste a lot of energy in communication among nodes and are not good fits for large scale applications. In the studies [12, 13], researchers studied an efficient-energy data collection method in connection with the scheduling policies of MAC. The author of the paper [12] defined a spatial correlation-based collaborative MAC protocol (CC-MAC) that regulated sensor node transmissions so as to minimize the number of reporting nodes. The author of the paper [16] considered the problem of correlated data gathering in sensor networks with multiple sink nodes and proposed a method with controlled sampling frequency. However, it had complicated process of optimization. The frameworks based on random sampling method [15, 18] were proposed to collect data in a more achievable way. The method [15] exploited packet overhearing to

choose the nodes for transporting data. Each node determined the redundancy of its reading by comparing the actual reading with the estimated value calculated by the overheard readings. Only the nodes whose data were required to grasp information on the whole area transmitted their data to sink. But the energy consumption for communication among nodes was too much at the beginning of each cycle. In the study [18], all nodes transmitted their readings to adjacent nodes with a fixed probability. The reporting nodes were the sampling nodes as “pacesetters”. The nonreporting nodes calculated the redundancy with the overheard data. If the difference between real value and the estimated value of a node was within a certain bound, it needed not to forward data to sink and went to sleep, otherwise it became a sampling node to transmit the readings. The algorithm chose reporting nodes reckoning without the residual energy of nodes and the number of iterations depends on the initial probability.

We proposed an energy-aware iterative method to choose sampling nodes to forward data. The probability of being chosen was related to the residual energy of nodes. The higher the energy was, the higher the probability of being selected was. So the life of the network could be prolonged because of the even energy consumption. In our algorithm, the initial probability could be a small number enough, and increased to 1 as the increase of iteration times. The size of the set of sampling nodes increased to the approximate optimal value at the last iteration. IDW (inverse distance weighting interpolation) is simple and a more appropriate interpolation method for the nodes with limited computing capability. Nodes used IDW to calculate the redundancy rate in our method.

3. Our Framework

Senor nodes usually are densely deployed in the monitoring area. In general, there is a high spatial correlation among the readings of adjacent nodes. So we can sample a part of nodes to transmit data to sink instead of all nodes. We proposed an Energy-aware Iterative Sampling Framework for data gathering in WSNs. The following is the detail introduces.

3.1. Redundancy determining

Inverse Distance Weighting interpolation method (IDW) is widely applied because of its ease [19, 20 and 21]. IDW is based on the close similarity principles: the shorter the distance between two things is, the more similar they are. The unknown measure at a position can be interpolated by means of a weighted average of the measures associated at the known points. In WSNs, IDW can be used to predict the unknown measure of a node with the known measures of the neighboring nodes. It is achieved by the following function:

$$\hat{V}_i^t = \frac{\sum_{j=1, j \neq i}^n \omega(i, j) V_j^t}{\sum_{j=1, j \neq i}^n \omega(i, j)} \quad (1)$$

\hat{V}_i^t stands for the interpolated value of node i at time t ; Node j belongs the set Z of adjacent nodes; $\omega(i, j)$ is the distance weight between node i and j . If $d(i, j)$ is the distance between node i and j , the weight can be expressed as followed:

$$\omega(i, j) = \frac{1}{d(i, j)^m} \quad (2)$$

The absolute difference between the predicted value \hat{V}_i^t and real measure V_i^t is the error ε called the redundancy of a node: $\varepsilon = |\hat{V}_i^t - V_i^t|$. Our purpose is to select a part of nodes as sampling nodes to forward data to sink with accuracy guarantees and put the redundant nodes into sleep model to save energy. Nodes decide whether they are redundant or not based on ε . The figure 1 shows the system overview, the black nodes are redundant nodes and the others are sampling nodes. The base station interpolates the global data and provides it to users.

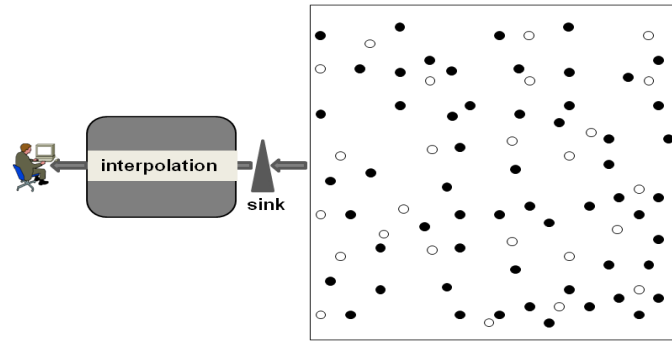


Figure 1. System overview

3.2. Energy-aware Iterative Sampling Algorithm

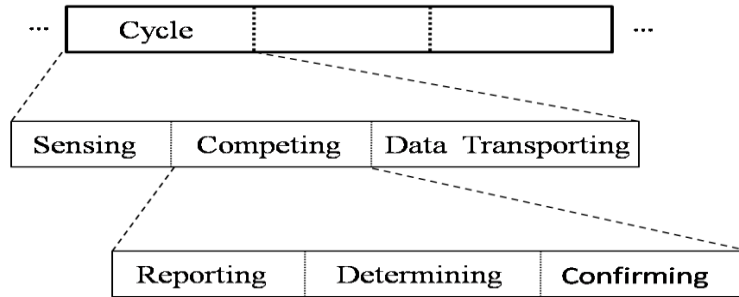


Figure 2. Overview of EISF

Figure 2 shows the overview of EISF. Each cycle consists of three phases, sensing phase, competing phase and data transporting phase. Competing phase consists of multiple TMEDA frames, reporting frame, determining frame and confirming frame. At first, all nodes keeping awake sense the data and then overhear the packets transmitted by the neighbors. In the competing frame, nodes will be chosen as reporting nodes with the individual probability. The probability is related with the remaining energy. We assume that the initial energy of a node is E_0 , the residual energy is E_r , $MaxRound$ (the number

of the most iterations) is k (k is more than 2), and the probability of the node at iteration i is:

$$P_i = \frac{CE_r}{E_0} (q)^{i-1} \quad (3)$$

P_i increases progressively with the iterations, and will increase to 1 at iteration $k-1$. So every node will have a chance of being a reporting node or not. And the selection of the set of sampling nodes (S) can be finished within certain iterative times k . We will get a sub-optimal set of sampling nodes at iteration k . $P_{k-1}=1$, we can get:

$$q = \left(\frac{E_0}{CE_r} \right)^{1/k-2} \quad (4)$$

Then the probability P_i is:

$$P_i = \left(\frac{CE_r}{E_0} \right)^{i-k+1/k-2} \quad (5)$$

Figure 3 shows the probability tendencies of the two nodes with different residual energy within ten iterations. The node with higher energy always has a higher probability in each iteration. The probabilities of the two node increase to 1 at ninth iteration.

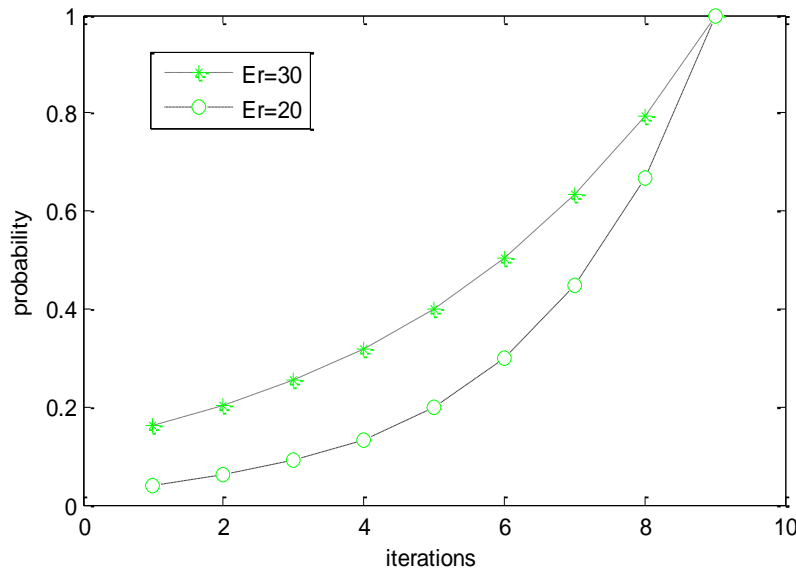


Figure 3. The probability tendencies

After each iteration, the sampling nodes broadcast their data packets to its neighbors with a certain transmit energy. A data packet consists of information on the identifier, the position and the readings of the transmitter. The nonsampling nodes overhear the data from the neighbors and calculate the individual redundancy according the information. If the redundancy isn't more than the error bound, the node will send conforming information consisting of its ID to the transmitters and then go to sleep until the next cycle, or not compete in the next iteration. Then the sampling nodes go to sleep until it is time to transmit data to sink at the end of the cycle. In the data transporting frame, all sampling nodes transport their data to sink. Algorithm 1 shows the energy-aware iterative sampling algorithm of nodes.

Algorithm 1:

```

1: All nodes sense the data at the beginning of a cycle
2: while Round  $\leq$  MaxRound
3:   for  $1 \leq i \leq n$   $\longleftarrow$  there are  $n$  nodes
4:     if node  $i$  isn't sleeping
5:       rand  $\longleftarrow$  a random number between 0 and 1
6:        $P_i = P(\text{Round})$   $\longleftarrow$  counts the probability
7:       if rand  $\leq P_i$ 
8:         put node  $i$  into  $S$   $\longleftarrow$  becomes a sampling node
9:         reporting its packet to its neighbors
10:      end
11:    end
12:  end
13:  for  $1 \leq i \leq n$ 
14:    if node  $i$  fail to be a sampling node
15:       $R$   $\longleftarrow$  determinesg redundant or not
16:      if  $R$  is true
17:        transmits confirming information
18:        goes to sleep  $\longleftarrow$  the reading the node is redundant
19:      end
20:    end
21:  end
22:  the sampling nodes go to sleep
23:  Round=Round+1
24: end

```

3.3. Gathering data

The nodes in the sampling set S transmit data to sink by multihop. Here we can use the geographical routing alike GEAR [22]. If a reading is sent by a sensor, it will be used. Then the sink interpolates the missing data by the same interpolation method based on the readings of the sampling nodes and the information about the ID of the nonreporting node in the confirming message. The sink will calculate the measure of one node with all readings consisting of its ID. When a node has no sampling data about its ID because of packet loss, its data can be interpolated with adjacent sampling nodes' readings. At last, all the data of the network will be collected within the error bound.

4. Performance Evaluation

4.1. Simulation Environment

The performance of our framework was evaluated using the energy model [23]. We compare EISF with ODSA [15] and E2K [18] in the same simulation environment. This real-world data refers to an Intel lab data set [24]. There were 54 nodes running for a month, monitoring temperature, voltage, humidity, and light. We populated the missing data with the average of the values during the previous and subsequent epochs at the same node. Our simulations were performed in Matlab 2009b using one-week temperature records.

4.2. Simulation Results

We proposed the energy-aware iterative sampling frame (EISF) to choose the sampling nodes set S . The number of the sampling nodes transmitting data to sink increases to an approximate optimal value while the iteration progressing. At last, more nodes save the energy of transmitting data. As the probability of being a sampling node has a relation with the current energy situation of a node, a node with higher energy has a higher probability to be a sampling node that balances the energy consumption. Other nodes estimate their readings based on the readings sent by sampling nodes and decide whether their readings are required to be transmitted or not. The nodes those are redundant or have very little energy will have more time to rest.

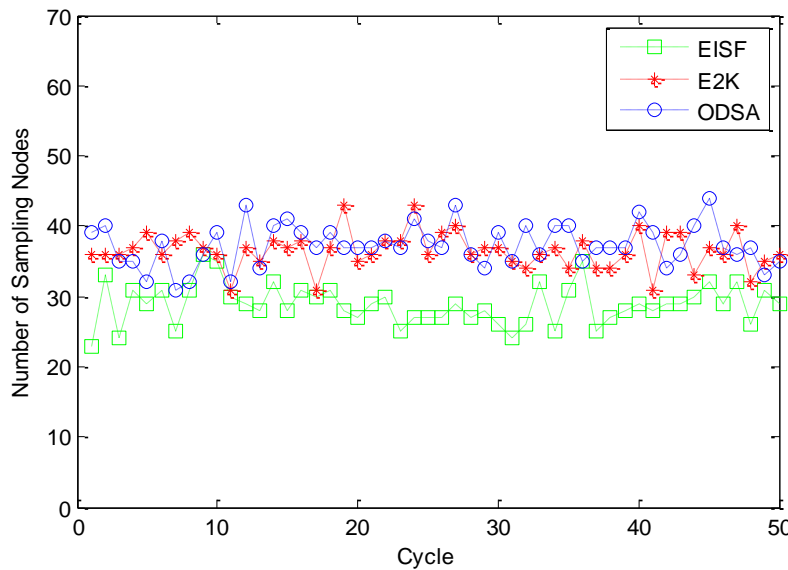


Figure 4. The number of the sampling nodes in every cycle

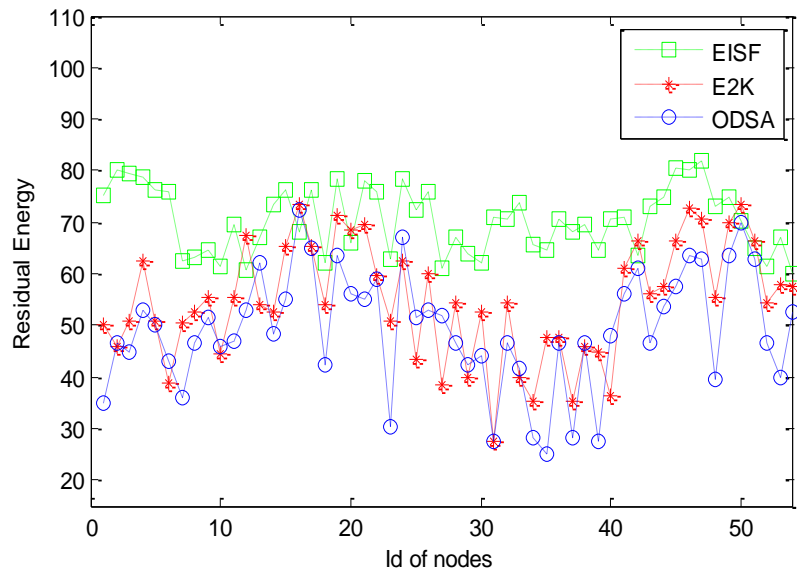


Figure 5. The residual energy of nodes in the 50th cycle

Figure 4 shows that the average number of the nodes of S in every cycle. EISF has a smaller size of the sampling nodes S in every cycle. In some cycles, the number is half of E2K's or ODSA's. So EISF could save more energy for transmitting data. Figure 5 shows the residual energy of the nodes in the last cycle. The residual energy of nodes was more than the other two methods. Our method had more effective energy consumption and made nodes die at the similar speed because of the energy-aware competing strategy.

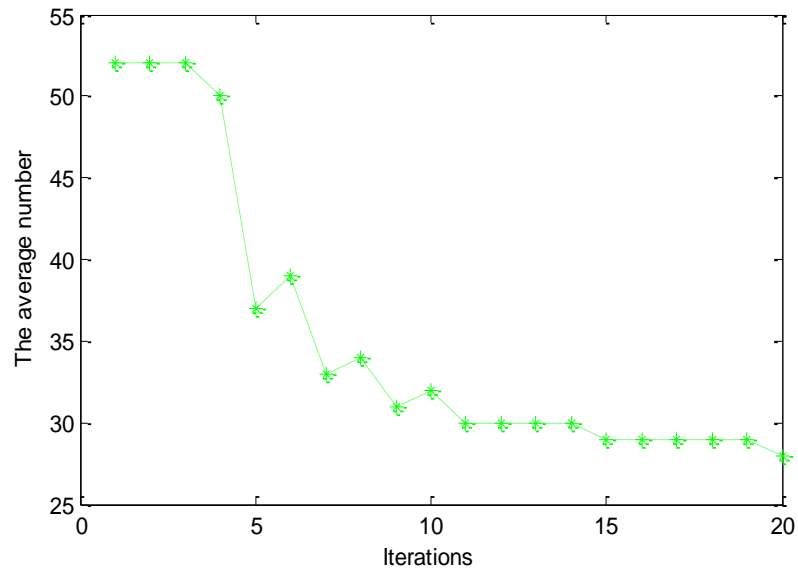


Figure 6. The number of the sampling nodes with different *MaxRound*

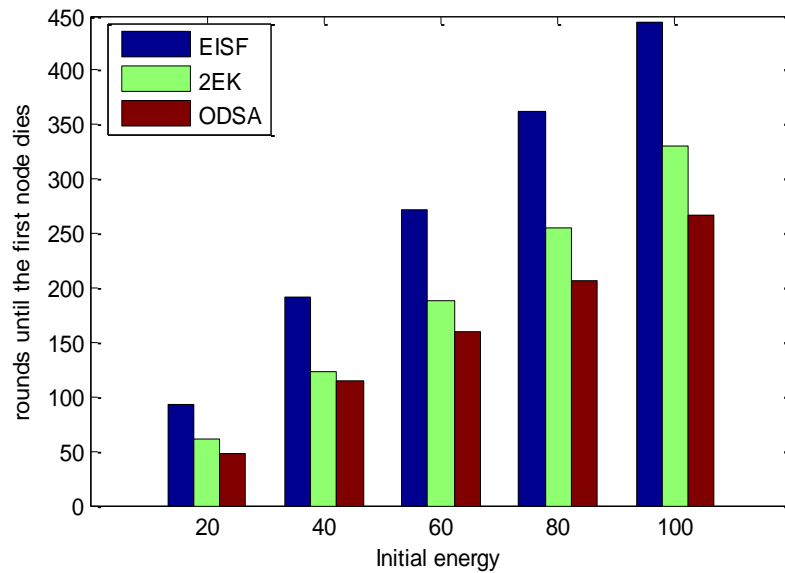


Figure 7. The life of the networks

The number of sampling nodes has an optimum for a deployed network. The result in the Figure 6 shows that the number is tending towards stability with the increasing of iterations. For the situation of the network of Intel lab, the number is stabilized after 10th iteration. So we can choose the sampling nodes set S within a certain number of iterations. As shown in Figure 7, our method prolonged the life of the network by 30%-50%.

5. Conclusions & Further Work

We have introduced our framework for gathering data: an energy-aware iterative sampling framework, choosing fewer nodes to finish the collection of data. The method reduced the communication traffic and prolonged the life of the network. The performance of our method was better than others' because of the energy-aware competitive iterative mechanism. In the further, we will explore the using of a new time-series forecasting model for reducing the temporal correlation of data for data gathering. We will apply the model to our framework.

Acknowledgments

The work in this paper has been supported by the National Natural Science Foundation of China (Grant No. 61005027), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 12KJB520009) and Scientific Research Fund of Nanjing University of Information & technology (Grant No. 20110442).

References

- [1] M. C. Vuran, O. B. Akan and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks", *Computer Networks*, vol. 45, no. 3, (2004), pp. 245-259.
- [2] M. Sharifzadeh and C. Shahabi, "Supporting Spatial Aggregation in Sensor Network Database", *ACM. GIS*, no. 4, (2004).

- [3] C. Liu, K. Wu and J. Pei, "An Energy-Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation", *IEEE Transactions On Parallel And Distributed Systems*, vol. 7, no. 18, **(2007)**, pp. 1010-1023.
- [4] A. Ali, A. Khelil, P. Szczytowski and S. Neeraj, "An Adaptive and Composite Spatio-Temporal Data Compression Approach for Wireless sensor networks", In *Proceedings of ACM MSWiM'11*, Miami, FL, USA, **(2011)**.
- [5] G. Anastasi, M. Conti, M. Di Francesco and A. Passarella, "Energy conservation in wireless sensor networks", *Ad Hoc Networks*, vol. 7, no. 3, **(2009)**, pp. 537-568.
- [6] A. Jain and J. Y. Chang, "Adaptive sampling for sensor networks", *Proc. 1st international workshop on Data management for sensor networks (DMSN 2004)*, Toronto, Canada, **(2004)** August 30th, pp. 10-16.
- [7] T. Arici and Y. Alrunbasak, "Adaptive Sensing for Environment Monitoring Using Wireless Sensor Networks", *Wireless Communications and Networking Conference*, **(2004)**, pp. 2347-2352.
- [8] R. Tynan, G. O'Hare, D. Marsh and D. O'Kane, "Intelligent agents for wireless sensor networks", In *Proceeding of The Fourth International Joint Conference on Autonomous Agents and Multi Agent System*, (AAMAS 05), IEEE Computer Society, **(2005)**.
- [9] M. Rahimi, M. Hansen, W. J. Kaiser, G. S. Sukhatme and D. Estrin, "Adaptive sampling for environmental field estimation using robotic sensors", in: *Proc. Of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, **(2005)** August 2-6, pp. 3692-3698.
- [10] J. Zhou and D. De Roure, "FloodNet: coupling adaptive sampling with energy aware routing in a flood warning system", *Journal of Computer Science and Technology*, vol. 22, no. 1, **(2007)** pp. 121-130.
- [11] A. Jindal and K. Psounis, "Modeling Spatially-Correlated Sensor Network Data", *Proc. Sensor and Ad Hoc Communication and Networks (SECON '04)*, **(2004)** October.
- [12] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks", *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, **(2006)**, pp. 316-329.
- [13] S. Chatterjea, T. Nieberg, N. Meratnia and P. J. M. Havinga, "A distributed and self-organizing scheduling algorithm for energy-efficient data aggregation in wireless sensor networks", *Technical Report TR-CTIT-07-10*, Enschede, **(2007)**.
- [14] J. M. Bahi, A. Makhoul and M. Medlej, "Data Aggregation for Periodic Sensor Networks Using Set Similarity Functions", *IWCMC*, **(2011)**.
- [15] Y. Iima, A. Kanzaki, T. Hara and S. Nishio, "An evaluation of overhearing-based data transmission reduction in wireless sensor networks", *Proc. Intl. Workshop on Sensor Networks Technologies for Information Explosion Era (Santa 2009)*, pp. 519-524.
- [16] K. Yuen, B. Liang and B. Li, "A distributed framework for correlated data gathering in sensor networks", *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, **(2008)**, pp. 578-593.
- [17] D. Chu, A. Deshpande, J. M. Hellerstein and W. Hong, "Approximate data collection in sensor networks using probabilistic models", In: *Proc 22nd International Conference on Data Engineering (ICD E06)*, Atlanta, GA, **(2006)** April 3-8, pp. 48-59.
- [18] B. Harrington, Y. Huang, J. Yang and X. Li, "Energy-Efficient Map Interpolation for Sensor Fields Using Kriging", *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, **(2009)**, pp. 622-635.
- [19] R. Tynan, G. M. P. O'Hare, D. Marsh and D. O'Kane, "Interpolation for Wireless Sensor Network Coverage", *IEEE EmNetS- II*, **(2005)**, pp. 123-132.
- [20] M. Vahabi and E. Tovar, "Data Gathering Approach in Dense Sensor Networks", **(2012)**.
- [21] P. Guccione, A. Ciampi, A. Appice, D. Malerba and A. Muolo, "Spatio-Temporal Reconstruction of UN-Sampled Data in a Sensor Network", In: *2nd International Workshop on Mining Ubiquitous and Social Environments*, **(2011)**.
- [22] Y. Yu, R. Govindan and Estrin, "Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks", *UCLA Computer Science Department Technical Report UCLA/CSD-TR-01-0023*, **(2001)** May.
- [23] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", *IEEE Trans. Wireless Comm.*, vol. 1, no. 4, **(2002)**, pp. 660-670.
- [24] <http://db.lcs.mit.edu/labdata.html>, **(2010)**.