AN ENERGY EFFICIENT SPATIAL CORRELATION BASED DATA GATHERING ALGORITHM FOR WIRELESS SENSOR NETWORKS

C.Tharini[1] and P. Vanaja Ranjan[2]

[1]Research Scholar, Department of Electrical and Electronics Engineering, Anna University, Chennai, India
`C_tharini@rediffmail.com`
[2]Department of Electrical & Electronics Engineering, Anna University, Chennai, India
`vranjanau@gmail.com`

ABSTRACT

Wireless Sensor Networks have a wide range of applications including environmental monitoring. These networks consist of wireless sensor nodes which are densely deployed to provide a wider coverage area. The dense deployment of the sensor node provides spatial correlation in the network. In this paper an efficient data gathering approach is implemented by combining the dual prediction and clustering algorithm. Clustering algorithm based on spatial correlation is used to cluster the sensor nodes. Then within the cluster, the nodes send their data to the sink using the Normalized Least Mean Square dual prediction algorithm. Simulation results show that the proposed algorithm reduces the average energy consumption of the network. .

KEYWORDS

Wireless Sensor Network, Prediction Algorithm, Clustering Algorithm

1. INTRODUCTION

Wireless Sensor Networks (WSNs) are an emerging field in wireless technology which requires deployment of several sensor nodes which communicate with each other to perform a common task. The sensor nodes are battery operated which is an important limitation of WSN. Several algorithms to conserve power are found in literature. These algorithms are designed to operate in various layers of the network to reduce power and to increase the lifetime of the sensor node.

In [1] it is stated that data transmissions can account for up to 70% of the power consumed in typical sensor node and substantial energy saving is possible if the volume of the communicated data is reduced using compression.

From [2] it is understood that clustering is a well established technique for reducing the data collection costs in wireless sensor networks. A survey of clustering algorithms is found in this paper. These algorithms vary in their objective which includes load balancing, fault tolerance, increased connectivity, reduced delay and maximal lifetime. In [3] sensor nodes are clustered and an efficient data gathering using wavelet based compression is implemented at the sensor node to reduce energy. A new collaborative data reduction method is proposed in [4] to remove the redundancy existing in the data coming from multiple sensors as well as from a single sensor. The proposed method uses a tree-based data propagation model to characterize the collaboration structure among multiple sensors. This algorithm uses PCA and partial correlation so that it can detect only linear data redundancy. Further studies are required to remove non-linear data redundancy.

In [5] the concept of data similarity is used to cluster the sensor nodes. This Data Similarity Variogram-based Clustering Algorithm (DSVCA) , is a good solution for network's data traffic minimization and for reducing the energy consumptions of the overall network. The algorithm is efficient when the sensor nodes are moderately correlated. In [6] spatial data correlations are exploited to group sensor nodes into clusters of high data aggregation efficiency. Temporal correlation in sensor data is not exploited in this work. In [7] major problems in applying Slepian-Wolf coding for data aggregation in cluster-based WSNs, including the clustered Slepian-Wolf coding problem and the optimal intra-cluster rate allocation problem is discussed. Hence Distributed Source Coding (DSC) requires an optimal rate allocation algorithm to perform an efficient aggregation. Distributed Source Coding algorithm which exploits spatial correlation in sensor data for data compression is proposed in [8], [9], [10]. These algorithms have not discussed the optimal rate allocation problem for DSC. In [11] clustering algorithms based on correlation is proposed but have not analyzed its effect on Distributed Source Coding or any other aggregation algorithms. In this paper a dual prediction algorithm is implemented along with the clustering algorithm and its performance in a clustered architecture is analyzed. The clusters are formed based on the spatial correlation in the network. In [14] an algorithm to approximate the time series instead of transmitting the data in raw form is proposed. This work focuses on reducing the communication by considering only the temporal correlation of the sensed data. In [15] a Distributed Clustering Algorithm (DCA) which is based on the use of predicted combined criteria metric is proposed. In this algorithm, the sensor with the highest predicted combined criteria in its r - hop neighborhood is selected as the cluster head. The next section describes the spatial correlation based data collection algorithm.

2. SPATIAL CORRELATION BASED DATA GATHERING ALGORITHM

The sensor nodes in the WSN continuously monitor the physical phenomena and communicate the observations to the sink. The sensor nodes tend to have high degree of spatial correlation as they are densely deployed in nature. Temporal Correlation also exists if the nodes sense physical phenomena like temperature humidity pressure etc. Hence there is a need for data reduction algorithms which exploits the correlation of data in sensor network. The proposed data gathering algorithm is designed for network with high spatial and temporal correlation.

The main objective of this algorithm is to reduce the communication between the source and the sink node. The algorithm works in two phases. In the first phase the sink node collects the data from all the sensor nodes. The sink node computes the magnitude similarity and trend similarity of the sensed data from the nodes. Then the nodes are partitioned into disjoint groups or clusters using the clustering algorithm. The clustering algorithm [12] partitions the sensor nodes based on the correlation metric. After obtaining disjoint group of sensor nodes the sink nodes provides the schedule with which the sensor node communicates to the sink node. In the second phase the sensor node transmits the data according to the schedule generated by the sink node. During data transmission a NLMS based prediction algorithm is performed synchronously at the source and the sink. The prediction algorithm predicts the value of the sensor node based on the previous values. If the predicted value is less than the predetermined threshold value then the sensor node will not communicate the sensed value. If the predicted value is more than the threshold then the sensor node communicates the current data to the sink node. The sink node instead of the predicted value it updates the received value and then continues the prediction with the received value. This prediction framework greatly reduces the transmission cost and hence reduces the energy consumption of the WSN. Figure 1 & 2 gives the function of the sink node and the source node
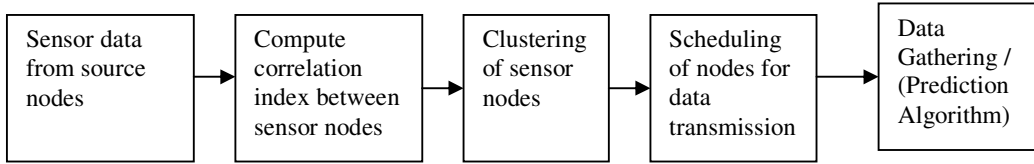
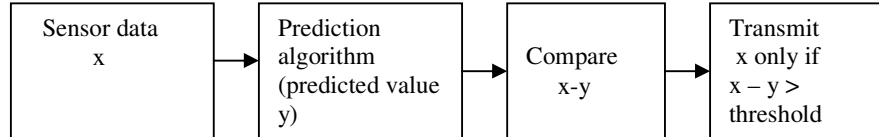| Sensor data from source nodes | → | Compute correlation index between sensor nodes | → | Clustering of sensor nodes | → | Scheduling of nodes for data transmission | → | Data Gathering / (Prediction Algorithm) |

Figure 1 Functions of Sink Node

| Sensor data x | → | Prediction algorithm (predicted value y) | → | Compare x-y | → | Transmit x only if x – y > threshold |

Figure 2 Functions of Source Node

The correlation metric and clustering algorithm used in this paper is from [12]. The magnitude dissimilarity and trend dissimilarity is used to find the correlation between the sensor nodes. The two parameters are defined as follows.

Magnitude dissimilarity: The two time series $X\{x_1, x_2, ..., x_n\}$ and $Y\{y_1, y_2, ...,y_n\}$ are magnitude m-similar if $(\sum^n_{i=1} | x_i - y_i |) / n <= m$

Trend dissimilarity: Two time series $X\{x_1, x_2, ..., x_n\}$ and $Y\{y_1, y_2, ..., y_n\}$ are trend t-dissimilar if $q1/q < t$,

where q1 is the total number of pairs (x_i,y_i) in the time series that satisfy $\Delta x_i \times \Delta y_i >= 0$, $\Delta x_i = x_{i-1}$, $\Delta y_i = y_{i-1}$, $1 <= I <= q – 1$.

2.1. Clustering Algorithm

The network is considered as a graph G with nodes N as the vertices. An edge is drawn between the node only if the dissimilarity index is less than or equal to the dissimilarity threshold value for intraclusters. This threshold is user defined because the correlation parameters can vary depending on the application. Once a graph G is constructed then the Clique partitioning algorithm is used to obtain disjoint cluster of the nodes. Figure 3 provides the steps used for partitioning the nodes into clusters. Once the cliques are formed, then every clique contains sensor nodes that are correlated with each other hence it is not necessary for all the sensor nodes within a cluster to transmit their information to the sink node. Therefore the sink performs a round robin scheduling wherein every node within a cluster will be transmitting one at a time. To further reduce energy a dual prediction algorithm is performed between sources and sink wherein the data is predicted both at the source and the sink using the past values from that node. Whenever the predicted value is more than the predetermined threshold then the sensor node transmits. This further reduces the energy consumed by the network. Normalized Least Mean Square Adaptive filter is used for predicting the data at both the source and the sink.

Step 1 Consider an Inputa graph G. The graph G has as many vertex as the number of nodes as the network

Step 2 The vertices of the graph are initially considered as uncovered.

Step 3 While all vertices are uncovered, a node with highest node degree is selected.

Step 4 Select all the vertices adjacent to v and put them into a list of S. The sensor node having the highest node degree is taken as V and its adjacent nodes are put in a new list S.

Step 5 A new graph Gtmp is generated consisting only the nodes in S.

Step 6 Tthe node degree of each vertex in Gtmp is obtained.

Step 7 Sort the vertices in S in ascending order of node degree in Gtmp;(To break a tie, the vertex with lower degree in the original graph G precedes);Construct a clique C containing only v;

Step 8 while (there are vertices available in S){ Pick up next vertex s from S; If s is adjacent to all vertices in C thus far, put s into the clique C;} . The adjacent nodes are combined with V to form a cluster having more than one node. The above process is repeated until all nodes come inside a Cluster.

Figure 3 Clique Partitioning Algorithm

2.2. Dual Prediction Algorithm

In the proposed algorithm a predictor runs concurrently at both the source and the sink to predict the sensed data based on the previous values of the data. As long as the predicted value remains within the threshold the sensed data is not communicated to the sink. Only when the predicted value exceeds the user defined threshold, then the data is transmitted. Normalized Least Mean Square (NLMS) algorithm is used for data prediction.

Normalized Least Mean Square Algorithm

The LMS algorithm is a type of adaptive filter known as stochastic gradient-based algorithms as it utilizes the gradient vector of the filter tap weights to converge on the optimal wiener solution. It is well known and widely used due to its computational simplicity. The equations (1), (2), (3) describe the algorithm

$$Y^j_{k.} = \Gamma_j^{(k)T} Z_{k,j} \qquad\qquad (1)$$

where $Z_{k,j}$ is the input vector, Γ_j is the weighting coefficient, Y^j_k is the predicted output

$$E_{k,j} = X^j_k - Y^j_k \qquad\qquad (2)$$

where $E_{k,j}$ is the error

$$\Gamma_j^{(k+1)} = \Gamma_j^{(k)} + \mu\, Z_{k,j}\, E_{k,j} \qquad\qquad (3)$$

μ is the step factor.

The step factor is assigned by trial and error in LMS algorithm In order to overcome this problem in normalized LMS the step size value is inversely proportional to the signal energy and is calculated using eq. 4.

$$\mu = 1/ (Z^T_{k,j} \, Z_{k,j}) + \Psi \hspace{4cm} (4)$$

where $Z_{k,j}$ is the input signal. The value of Ψ is a small positive constant in order to avoid division by zero when the values of the input vector is 0.

2.3 Results

The proposed algorithm is implemented in Matlab Wireless Simulator. The following are the parameters considered for the simulation.

Simulator	**Wireless Matlab Simulator**
Total area	**1600 meter square**
Simulation time	**5s**
Sink	**Static Sink**
Topology	**Single Hop**

All the nodes transmit same amount of data to the sink. Simulation is performed to calculate the total energy consumption of the network under two conditions. First the simulation is performed with normal sensor network and the simulation is performed using the proposed algorithm. Average energy consumption of the network is computed using the energy model as given in [11]. The transmitter power consumption and receiver power consumption is obtained from the datasheet of PIC16F877A processor and is considered as 63mW and 61.8 mW respectively. From the results graph it is observed that the proposed algorithm reduces a greater amount of communication required in transmission of the sensed data. Initially all the sensors transmit their values to the sink. This enables the sink to compute the spatial correlation between the nodes. Clusters are formed based on the spatial correlation. Then within each cluster the nodes transmit their data using dual prediction algorithm.
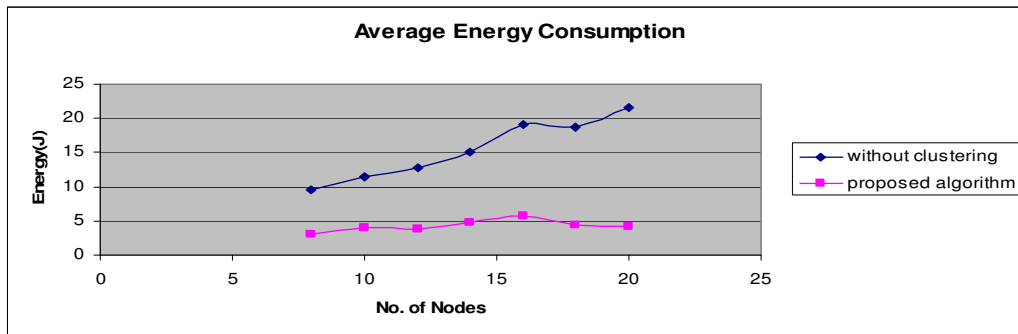

Figure 4 Average Energy Consumption after clustering the nodes

Figure 4 shows the average energy consumption for a network with different number of nodes. This algorithm was implemented for the temperature data. This clique partitioning algorithm groups sensors into different clusters. The nodes in each cluster transmit their data using prediction algorithm. The nodes transmit only when the predicted value exceeds the threshold value. This in turn reduces the average energy consumption of the sensor nodes. Figure 5 provides the simulation output for transmission of data to sink using wireless sensor network simulator in matlab. Figure 6 provides the simulation output when the difference exceeds the given threshold, then the source node transmits the corresponding sensor data. This value will be updated by the sink.
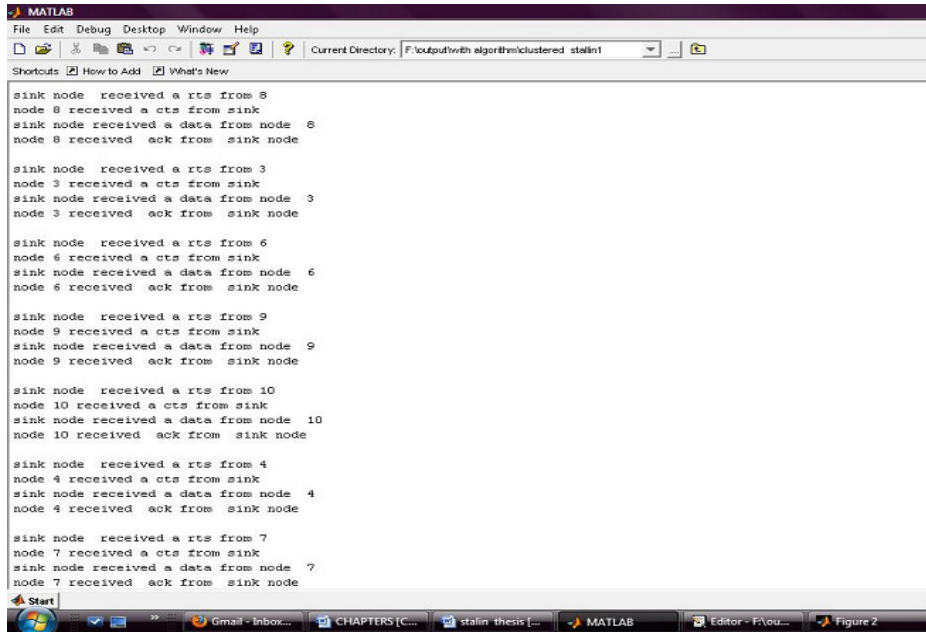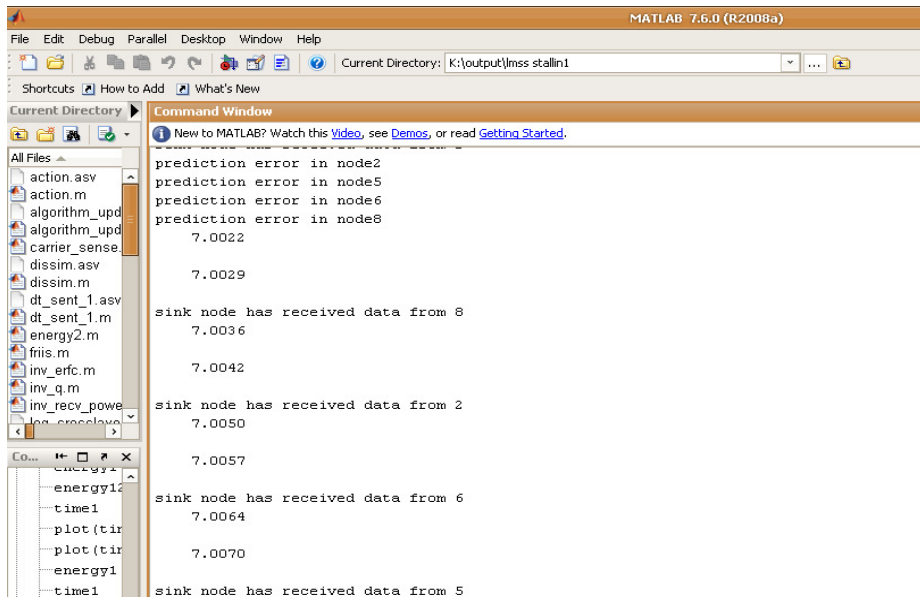
Figure 5 Transmission of Data to Sink
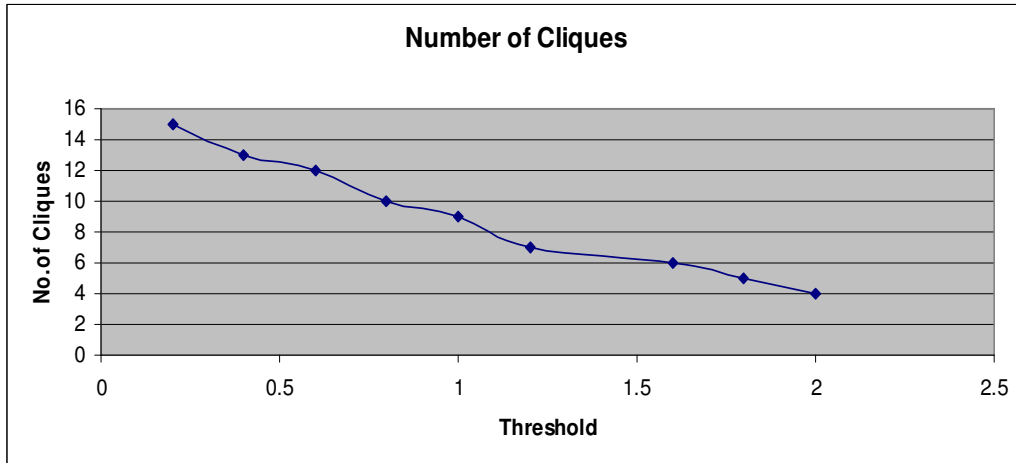

Figure 6 Dual prediction

Figure 7 Number of Cliques for different magnitude dissimilarity threshold value

Figure 7 provides the number of cliques formed when the dissimilarity threshold is varied. The number of transmission shall be reduced only if the number of cliques is less. But it is observed that when the threshold value is increased then the number of cliques formed is reduced. Therefore there is a trade off in the accuracy of the predicted data and the number of transmissions in a network.
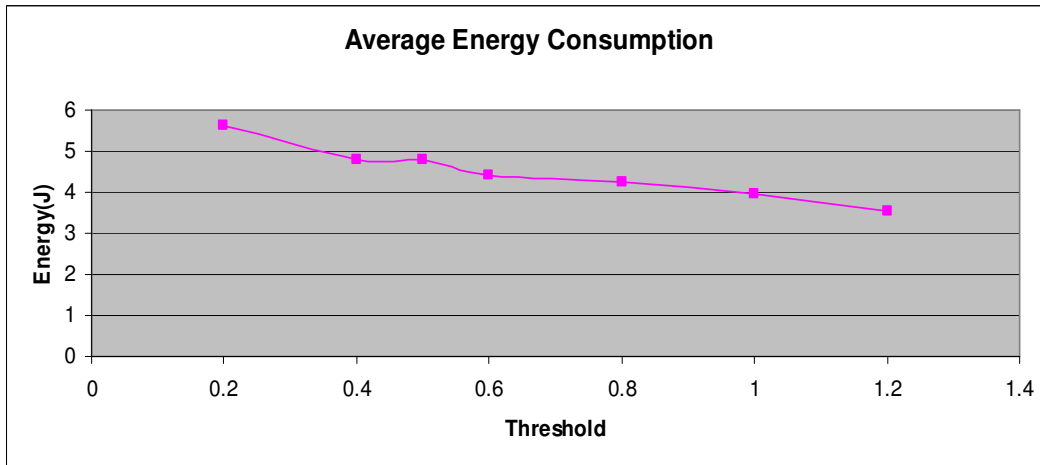

Figure 8 Average Energy Consumption for different threshold values

The average energy consumption of a network with 20 nodes when the dissimilarity value is varied is shown in figure 8. It is obvious that only minimum number of cliques is formed when the dissimilarity calculation threshold is high. Therefore there are less number of transmission as the nodes transmit their data based on the scheduling generated by the sink node.

3. CONCLUSIONS

It is observed from the simulation results that the proposed algorithm is efficient when there is high spatial and temporal correlation. Change in correlation characteristics is adapted by periodical data collection from all the sensor nodes. The threshold value used in the prediction algorithm is user defined and it depends on the application. NLMS predictor is used in the proposed algorithm. This is suitable for any input signal. Future work may involve newer predictors for fast transient signals. The proposed algorithm is energy efficient and the results show 29% reduction in average energy consumption when compared with a network without clustering the nodes. Future work will include the evaluation of the performance of network by considering different clustering algorithms and also by considering different predictors. Also clustering algorithm which adapts to changes in the correlation characteristics can be considered in future.

4. REFERENCES

[1]   H.Cam, S Ozdemia, P. Nair, D. Muthuavinashiappan, H. Ozgur Sanli, (2006) "Energy-Efficient Secure Pattern based Data Aggregation for Wireless Sensor Networks", *Journal of Computer Communications*, vol.29, no. 4, pp 446 – 455.

[2]   A. Abbasi and M.Younis, (2007) " A survey on Clustering Algorithms for Wireless Sensor Networks", *Computer Communications*, vol 30, no. 14 – 15, pp 2826 – 2841.

[3]   Qianqian Ren, Jianzhong Li, Jinbao Li, (2007) "An Efficient Clustering-Based Method for Data Gathering and Compressing in Sensor Networks", *International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, pp 823-828.

[4]   Chiwoo Park, Yu Ding, Eunshin Byon, (2008) "Collaborative Data Reduction for Energy Efficient Sensor Networks", *4th IEEE Conference on Automation Science and Engineering*, pp 442 – 447, August 23-26.

[5]   Laura Imbriglio, Fabio Graziosi, (2009) **"**DSVCA: A Novel Distributed Clustering Algorithm for Wireless Sensor Networks based on Statistical Data Correlation", *International Conference on Ultramodern Telecommunication and Workshop*, pp 1-6.

[6]   Chongqing Zhang, Binguo Wang, Sheng Fang, Jiye Zheng, (2008) "Spatial Data Correlation Based Clustering Algorithms for Wireless Sensor Networks", *Intetnational Conference on Innovative Computing Information and Control (ICICIC'08)*, pp 593.

[7]   Pu Wang and Cheng Li, Jun Zheng,(2007) "Distributed Data Aggregation Using Clustered Slepian-Wolf Coding in Wireless Sensor Networks", *International Conference on Communications*, pp 3616 – 3622.

[8]   Caimu Tang, Cauligi S Raghavendra and Viktor K Prasanna, (2003) "An Energy Efficient Adaptive Distributed Source Coding Scheme in Wireless Sensor Networks', *International Conference on Communications*, pp 732 – 737.

[9]   Guogang Hua and Chang Wen Chen, (2005) "Distributed Source Coding in Wireless Sensor Networks", *International Conference on Quality of Service in Heterogeneous Wired and Wireless Networks*, pp 6.

[10]   Jim Chou and Kannan Ramachandran, (2003) "A Distributed and Adaptive Signal Processing Approach to Reducing Energy Consumption in Sensor Networks', *International Conference on IEEE Computer and Communication*, vol. 2, pp 1054 – 1062.

[11]   S. Besagni, (1999) "Distributed Clustering Algorithm for Adhoc Networks', *In proc. of the International Symposium on Parallel Architectures, algorithms and Network*, pp 310 – 315.

[12]   Chong Liu, Kui Wu, and Jian Pei, (2007) "An Energy Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation", *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, issue 7, pp 1010-1023.

[13]   Wang A and Chandrakasan A, (2002) "Energy-Efficient DSP's for Wireless Sensor Network". *Signal Processing Magazine*, vol. 19, Issue 4, pp 68-78.

[14]   Ngcoc Duy Pham, Trong Duc Lo, Hyunseung Choo, (2008) "Enhance Exploring Temporal Correlation for Data Collection in WSN's", *IEEE International Conference on Research, Innovation and Vision for the future*, pp 204 – 208.

[15]   Koucheryavy, A, Salim, A, (2010) "Prediction-based Clustering Algorithm for Mobile Wireless Sensor Networks", *International Conference on Advanced Communication Technology*, pp1209 – 1215.