ISSN: 0976-3104
SPECIAL ISSUE (ASCB)
Rajesh et al.

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

DATA AGGREGATION IN NOISY WIRELESS SENSOR NETWORKS USING CHAOS THEORY

G. Rajesh*, **S Mathivanan, B. Vinayagasundaram**

Dept. of Computer Science, Anna University, Tamil Nadu, INDIA

ABSTRACT

In environmental monitoring applications, the data periodically sensed by Wireless sensor networks (WSNs) have a strong redundancy. The Network consumes more power to transmit the redundant data packets to sink through intermediate nodes which reduces network lifetime. To overcome the redundant data transmission, Chaos Theory based Data Aggregation (CTAg) prediction method is proposed here. The proposed method minimizes the number of packets forwarded to sink node by eliminating redundant packets by using chaos theory based prediction technique. The proposed approach is evaluated using temperature monitoring application dataset collected from Intel Berkley Lab. The CTAg method significantly reduces communication redundancy, data redundancy, and mean square deviation error and also increase prediction accuracy, in turn evidently proved that the lifetime of the network is improved.

Received on: 18th-March-2015
Revised on: 20th-May-2015
Accepted on: 26th– June-2015
Published on: 8th –Aug-2015

KEY WORDS

Wireless sensor network; Redundancy; Data aggregation; Time series prediction; Chaos theory;

***Corresponding author: Email:** raajiimegce@gmail.com, gr@annauniv.edu **Tel:** +91-044-22516023

INTRODUCTION

Wireless sensor networks (WSNs) are groups of sensor nodes that are tiny with low power and low cost. Each node consists of a sensing unit, processing unit and communication subsystem [1]. These sensor nodes collect real-time data from the physical environment and transfer the collected data to base station. They are widely used in military, health science, and commercial applications. For example, in health science, a doctor can monitor the physiological information about the patients remotely. In the commercial application, it is used for managing inventory and monitoring product quality [2].

The energy consumption in a wireless sensor network is due to sensing, computing, communication and mobility of nodes. The energy for communication is more expensive, which affects the energy of intermediate nodes during multi hop transmissions of data packets compared to sensing and computation. Energy dissipation during communication that is caused by transmitter or receiver is considered as E_{elec} and the power dissipation of the transmitter amplifier is taken as ε_{amp}. The power consumed by the transmitter for k-bit packet transmission [3, 6] to a distance‘d’ and assume d^m as path loss.

$$E_{Tx}(k, d) = E_{elec} * k + \varepsilon_{amp} * d^m * k \qquad (1)$$

Let $E_{elec} = E_T(k, d) = E_R(k) = 50nJ/bit$, ε_{amp}=100pJ/bit/m^2 and $m = 2$.

$$E_{Rx}(k) = E_{elec} * k \qquad (2)$$

Where $E_{Rx}(k)$ is the k–bit message in receiving of energy.

In an environmental monitoring application, information collected by sensor nodes tends to exhibit strong temporal correlation. This redundancy leads to communication overhead, consumes node energy and reduces network efficiency. To overcome this issue, the data aggregation techniques [5] were introduced. Wireless sensor networks are partitioned into source node, intermediate node

ISSN: 0976-3104

(aggregator node) and sink node. The aggregator node performs the data aggregation [4]. It collects data from multiple sensor nodes, which are fused and transferred to the base station. Thus the data aggregation technique eliminates redundancy and minimizes numerous transmissions and thus preserves the network energy.

Chaos theory is the study of complex, nonlinear, dynamic systems [16]. It is a branch of mathematics that deals with systems that appear to be orderly (deterministic) but, in fact, harbor chaotic behaviors. It also deals with systems that appear to be chaotic, but, in fact, have underlying order. Chaos theory studies the behavior of dynamical systems that are highly sensitive to initial conditions, an effect which is popularly referred to as the butterfly effect. The deterministic nature of these systems does not make them predictable. This behavior is known as deterministic chaos, or simply chaos.

The proposed Chaos Theory based Data Aggregation (CTAg) approach predicts the next period data based on the earlier sensed data. The data are collected from ordinary sensor nodes. The aggregator node made a decision between prediction error and prediction threshold, in order to make a conclusion whether to transmit current data to sink node. This technique eliminates the redundancy and minimizes the communication density. The proposed method can also provide better prediction accuracy and prolong sensor node lifetime when compared to other traditional prediction based data fusion techniques. This paper is organized as follows: Section II discusses the prediction base data aggregation and chaos theory based data aggregation (CTAg). Moreover, implementation details and results are given in Section III. Finally, conclude and define our future work in Section IV.

MATERIALS AND METHODS

Prediction based data aggregation- a survey

The energy management is one of the major issues in wireless sensor networks. A sensor utilizes high energy for communication rather than sensing and processing. The redundant communication in noisy channels causes the depletion of network energy. The prediction based data aggregation approach reduced unnecessary data transmission and so energy expenditure in communication subsystem was minimized. Hyuntea Kim et al., [8] exploited linear data prediction method to improve communication efficiency and to minimize energy consumption with data correlation. As the model is designed considering some factors such as the selective transmission, it reduced data accuracy and adjustments in aggregation period caused the network to meet the additional delay. Guiyi Wei et al., [7] proposed a method that saves network energy and eliminates redundant communication by exploiting prediction based data aggregation protocol. However, in this method synchronization time increased due to synchronization has to be done prior to each transmission. Guorui Li et al., [9] proposed an Auto Regressive Integrated Moving Average Model (ARIMA) that predicts the next time value based on the previous observed values. When the prediction error is less than the preconfigured threshold value the aggregator would not transmit the data sensed by the source node. Otherwise, it transmits the data to sink node. Therefore ARIMA model reduced the amount of data transmitted between the ordinary sensor node and aggregator node. Since this method performed aggregation on the ordinary sensor node and aggregator node it increased the computational complexity and reduced accuracy. Rajesh G et al., [10] proposed the data fusion method using Simpson's 3/8 rule to forecast next time data based on the early sensed information. When prediction error is greater than the prediction threshold the cluster head transmits the actual sensed value to the base station. Otherwise, it would not transmit data to the base station. This method reduced unnecessary transmission between cluster head and base station. However, this method provides less prediction accuracy since the deviation error is increased between subsequent values. There are several data fusion techniques in Wireless sensor networks. The main features of the proposed work are that it, Improves the performance of the forecast and Performs less computation to obtain the forecasted data.

Chaos theory based data aggregation (CTAg) technique

The typical features of chaos include: 1) Nonlinearity. If it is linear, it cannot be chaotic. 2) Determinism. It has deterministic underlying rules every future state of the system must follow. 3) Sensitivity to initial conditions. Small changes in its initial state can lead to radically different behavior in its final state. Long-term prediction is mostly impossible due to sensitivity to initial conditions. A dynamic system is a simplified model for the time-varying behavior of an actual system [17]. These systems are described using differential equations specifying the rates of change for each variable. A dynamical system of dimension N system first–order differential equations for N variables $x_1(t), x_1(t) \dots x_N(t)$ evolve with time t according to,

$$\hat{x}_1 = f_1(x_1, x_2, \dots x_N, t) \quad (3)$$

$$\hat{x}_2 = f_2(x_1, x_2, \dots x_N, t) \quad (4)$$

$$\hat{x}_N = f_N(x_1, x_2, \dots x_N, t) \quad (5)$$

Where f_1, f_2 are assigned functions and a dot is a derivative with respect to time.

THE IIOAB JOURNAL ISSN: 0976-3104

The system following Characteristics of a Chaotic System:

- Sensitivity to initial conditions
- Non-linear
- Dynamic and mixed topology system and Continuous or periodic time.

So that the Chaos is the aperiodic long–term behavior in a deterministic system that exhibits sensitive dependence on the initial condition. These characteristics enables chaos theory based data aggregation (CTAg) prediction method is suitable for eliminating data redundancy in WSNs.

Considered a hierarchical wireless sensor network G (SN, E) where, SN represents the sensor nodes and E represents links connecting the nodes. These sensor nodes collect weather monitoring data (Temperature, Humidity) periodically. Each node transmits data to sink node through the intermediate node or aggregator node (A). The aggregator (A) will perform data fusion by eliminating redundant data using chaos theory before transmitting the gathered data towards the base station. This will minimize the amount of data transmitted between aggregator node and sink node.

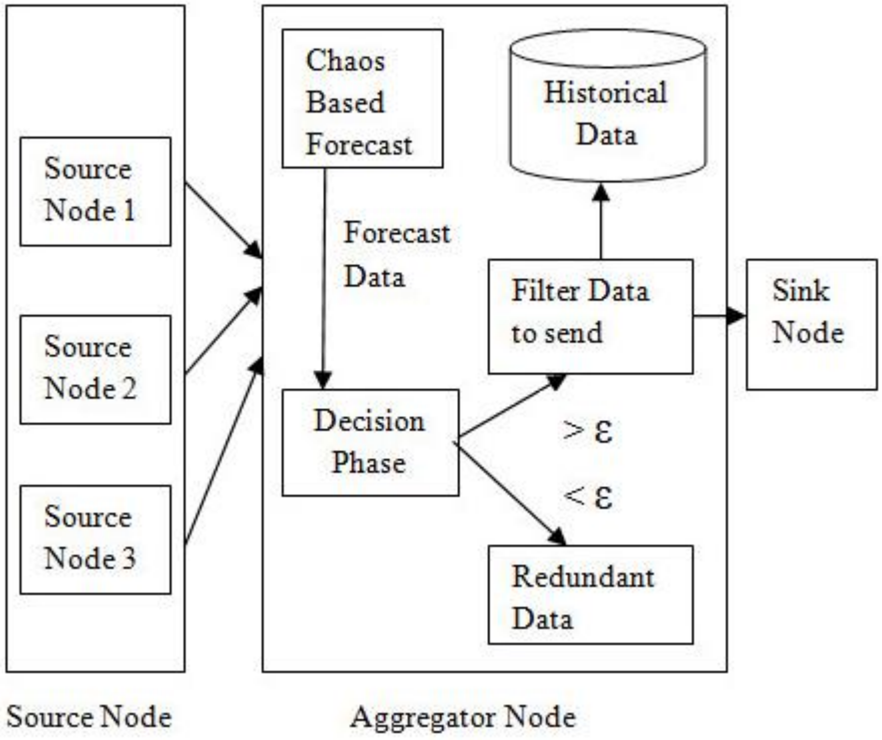

Fig: 1. Architecture for chaos theory based aggregation.

Figure–1 illustrates, the CTAg architecture that consist of two distinct phases, chaos based forecasting and decision phase. The chaos based forecasting phase uses previous time series measurements representing (6) to compute next time period data using (7) and (8). The decision phase will accomplish the comparison between forecast errors e_t and prediction threshold ε to make a decision whether to send actual sensed data to the sink or not. The prediction error was calculated using (10). The prediction error is greater than the prediction threshold the aggregator node transmits sensed data to base station. The aggregator does not send sensed data to sink node when forecast error is less than the prediction threshold.

$$x_t = \left(x(t_n), x(t_{n+\tau}), x(t_{n+2\tau}), \ldots, x\left(t_{n+(m-1)\tau}\right)\right) \qquad (6)$$

Where x_t is the chaotic time series, m is the embedding dimension and τ is the delay time.

$$x'_{t+n} = f(x) \qquad (7)$$

Where x'_{t+n} the prediction data of next time and n is step length of time series forecast.

$$f(x) = x_t + \left(\tau - (\tau - n)\right) * FE \qquad (8)$$

THE IIOAB JOURNAL ISSN: 0976-3104

Where, $\tau = \frac{x - x_{n-t}}{h}$ x is the recent data, x_{n-t} is the previous time data and $h = t - t_n$. Let be t is the time for recent data sensed and t_n is the time for earlier sensed data value.

The Forward Error (FE) Δ is the product of Minimum Forward Error Value (MFEV) with current sensed data and divide to embedded dimension is the order of MFEV.

$$FE\ (\Delta) = \frac{\Delta^m x_t}{m!} \qquad (9)$$

$$e_t = x_{t+1} - x'_{t+1} \qquad (10)$$

Where e_t is the prediction error at period t.

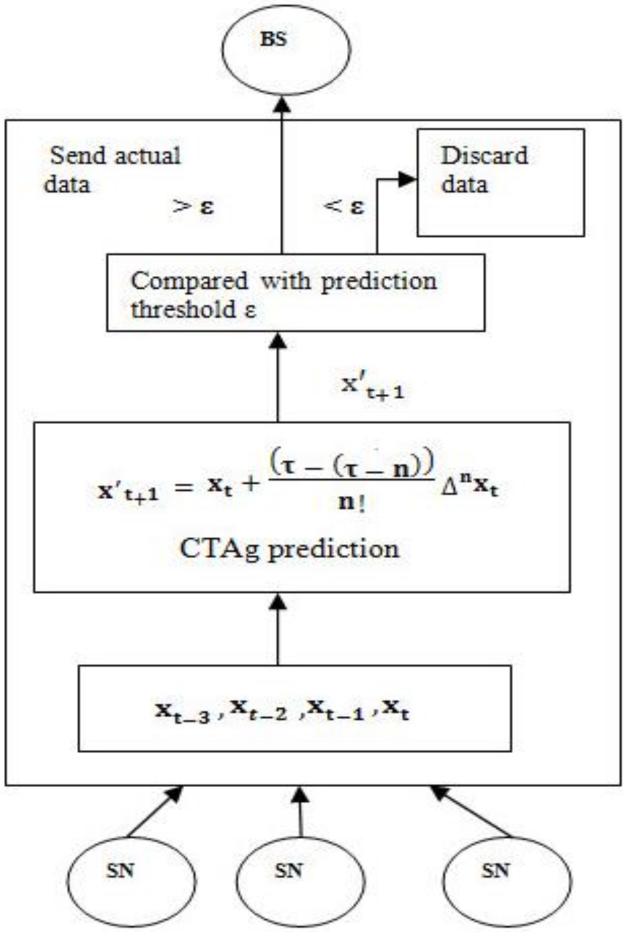

Fig: 2. Chaos time series forecasting model.

Figure–2 illustrates, when an intermediate node performs chaotic time series forecasting is evaluated future value x'_{t+1} based on the Sensor Nodes (SN) gathered previous time temperature data such as $x_{t-3}, x_{t-2}, x_{t-1}, x_t$... after computing predicted value, to measure the forecast error $x_{t+1} - x'_{t+1}$ mean that the deviation between actual sensed value and prediction value, then it is compared with prediction threshold value ε. If the prediction error is higher than the threshold, the actual sensed data transmitted to sink node. Otherwise, the aggregator is not forwarding the packet to the Base Station (BS). Here the aggregator node, predicted threshold based will perform filtering the redundant data. The predicted threshold set as the mean value of the prediction error. The predicted threshold is adaptive. The less deviated data is discarded from gathered continuous time series data set. The more relevant data or non-redundant data is towards base station (BS). But the major and most significant limitation of

THE IIOAB JOURNAL
ISSN: 0976-3104

chaos theory is the feature that defines it, sensitive dependence on initial conditions that may affect prediction accuracy.

Algorithm 2, illustrates the aggregator node receives predicted data at the point of time $t+1$. The aggregator decides to transmit data towards the sink or not by doing a comparison with predicted error and prediction threshold.

Algorithm 1: Chaos Forecasting Phase

Input:

x_t → Time series sensor data

Output:

x'_{t+1} → Forecasted Data

Process:

1. Get earlier time ser $\Delta^m x_t$

x_t = {x(0), x(1) m2),x(3),..,x(t)}

2. If History_size > 0

2.1. Compute $\Delta x_t = x_t - x_{t-1}$

2.2. Compute the next time series of the data

$$x'_{t+1} = x_t + \frac{(t-(t-n))}{n!} \Delta^n x_t$$

2.3. Return x'_{t+1}

3. End if

Algorithm 2: Decision Phase

Input:

x'_{t+1} Forecasted Data

x_{t+1} Actual Sensed Data

Output:

Returned Actual Sensed Data

Process:

1. At the period t aggregator node computes x'_{t+1}

2. At period t+1, aggregator node receiving actual data x_{t+1}

3. If $(x_{t+1} - x'_{t+1} < \varepsilon)$

3.1. Redundant Data

3.2. [History] = Sensed data

4. Else

4.1. Transmit Non-Redundant Data

4.2. [History] = Forecast data

5. End If

THE IIOAB JOURNAL
ISSN: 0976-3104

If the prediction error variation is less than a threshold value the aggregator assumes that the sensed data is redundant data. So it does not forward to sink node. Otherwise, assume that the non-redundant data and it conveyed to sink node.

RESULTS

The CTAg algorithm simulated using OMNET++ simulator with MIXIM package. The actual time series data are collected from wireless sensor networks. Here N=100 nodes are uniformly used to collect data from a sensing area of $A = 500 \times 500\ m^2$. The data are traversed using AODV protocol. In this network, the sensor node spends E_{Tx} energy to carry the data 'd' distance to next node and utilizes energy E_{Tr} for receiving data.The total communication energy is e_i. The CTAg is intended to reduce redundant communication and conserve the network energy e_i. In this section, the CTAg algorithm is applied to the temperature monitoring application. The dataset is collected from Intel Berkley Lab [17].

Table–1 represents the table contains date, time, and node id of the sensed data. The Epoch is a monotonically increasing sequence number from each node. In this dataset we consider the Time (t), Mote id (MID), Epoch (E) and Temperature (T) for evaluation. The CTAg algorithm performance is compared with conventional method of Kalman Filter [KF] prediction based data aggregation [15]. It is observed that the proposed algorithm is better among the data aggregation technique.

Table: 1. Example temperature application dataset

Date	Time	Epoch	Mote id	Temperature
03/03/04	09:45	22	1	20.145
03/03/04	10:22	23	1	20.165
03/03/04	10:46	24	1	20.165
03/03/04	11:16	25	1	20.169
03/03/04	11:20	26	1	20.244

Performance analysis

Table–2 represents the comparison of quantitative performance of proposed CTAg algorithm and KF. This table contains the actual data, predicted data, prediction error, and data transmission status and prediction threshold. Here, the prediction threshold as 0.28 is used to evaluate the performance of prediction. In Data Transmission Status, 0 indicates that the data is similar to the previously sent data and 1 indicates that the data is different from previous data. The aggregator will transmit to the sink node data marked as 1 and discard data marked as 0. Hence the communication overhead is reduced. The communication overhead is the ratio of the sensor node actual power consumes the number of packet forwarded to base station (BS) to power consumes after aggregated packet transmitted to base station (BS). Thus the CTAg aggregation method reduces the redundant packet transmission and power consumption.

Figure–3 represents comparison of actual sensed temperature with Kalman Filter [KF] and CTAg method forecasted the temperature (T) in different time interval. The temperature predicted using CTAg method is close to actual sensed value due to small prediction error and KF method has high variability with sensed data and CTAg due to high deviation error between the sensed and predicted values. It can also be seen that the predicted value of CTAg varies consistently with the observed value and forms a steady growth curve with almost constant prediction error. This stable prediction error rate eliminates the need for the fixing prediction threshold using soft computing technique.

THE IIOAB JOURNAL ISSN: 0976-3104

Figure–4 shows the amount of deviating data that has been transferred. The CTAg method performs better for the forecast threshold up to 0.28 when compared to KF. The CTAg restricts redundant packet transmission from aggregator to sink node compared to the Kalman filter.

Table: 2. Sample resultant data

Method	Actual Data	Predicted Data	Prediction Error (%)	Data Transmission Status
CTAg	20.51	20.77	0.26	0
	20.54	20.79	0.25	0
	20.61	20.92	0.32	1
	20.86	21.04	0.18	0
KF	20.51	20.81	0.30	1
	20.54	20.78	0.24	0
	20.61	20.92	0.31	1
	20.86	21.13	0.27	0

Figure–5 based on the forecast threshold 0.03, 0.08, 0.13, 0.18, 0.23, and 0.28. When the predicted threshold is 0.28 the difference between the observed and forecasted value is minimum. Hence it can be inferred that considerable accuracy is achieved at prediction threshold 0.28.

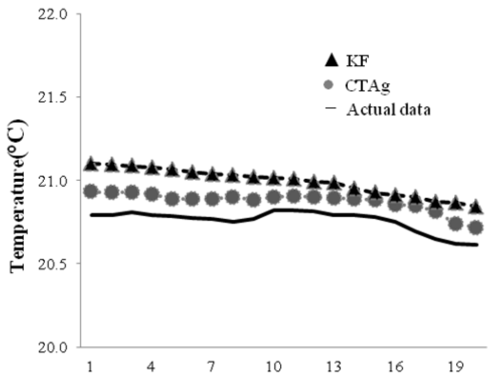

Fig. 3

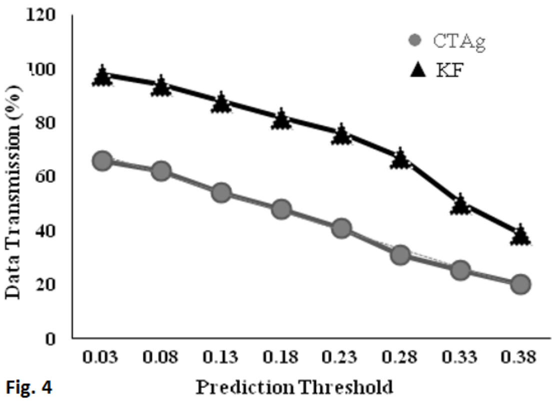

Fig. 4

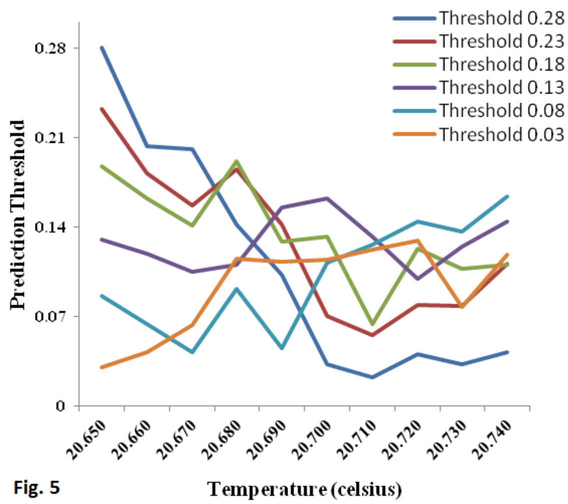

Fig. 5

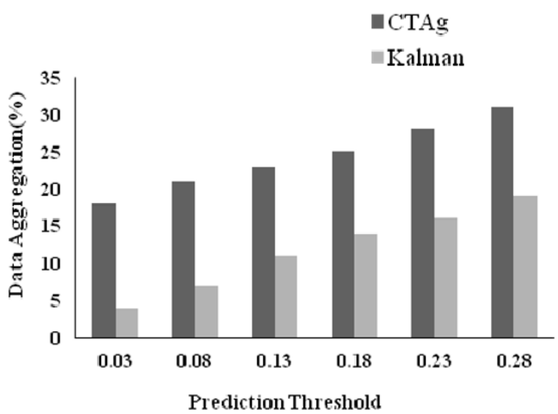

Fig. 6

Fig: 3. Comparison of prediction and sensed temperature in time. Fig: 4. Comparison of data transmission (%) in variable temperature. Fig: 5. Comparison of adaptive threshold. Fig: 6. Comparison of data aggregation (%) in different threshold.

In **[Figure–6]**, based on the forecast threshold, the number of data aggregation is larger when compared to the Kalman filter. Due to this data aggregation is reduced redundant transmission and communication consumption power.

CONCLUTION

The energy efficiency is the important key Wireless sensor networks. With data transmission is the major part of energy consumption, chaos theory based time series prediction method to enhance energy efficiency. The proposed Chaos Theory based Data Aggregation (CTAg) based approach reduces redundant data, communication overhead and number of packet transmission between aggregator and sink node by using adaptive thresholds. The time series prediction using CTAg method was energy efficient and performed less computation to obtain the forecasted data. The experiments also show CTAg achieves better performance compared to other prediction approaches like Kalman Filter [KF] based prediction.

.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

ACKNOWLEDGEMENT

None.

FINANCIAL DISCLOSURE

No financial support was received to carry out this project.

REFERENCES

[1] Akyildiz IF, Su W, Sankarasubramaniam Y, Cyirci E. [2002] Wireless sensor networks: A survey. *IEEE Communications Magazine*: *August*.

[2] Yick J, Mukherjee B, Ghosal D. [2008] Wireless sensor network survey. *The International Journal of Computer and Telecommunications Networking*. dl.acm.org/citation.cfm?id=1389832

[3] Heinzelman WR, Chandrakasan A, Balakrishnan H. [2000] Energy-Efficient Communication Protocol for Wireless Microsensor Networks. *Proceedings of the 33rd Hawaii International Conference on System Sciences*

[4] Krishnamachari B, Estrn D, Wicker S. [2002] The impact of data aggregation in wireless sensor networks. *Proceedings of International conference on Distributed Computing Systems Workshops (ICDCSW).*

[5] Rajagopalan R, Varshney PK. [2006] Data-aggregation techniques in sensor networks survey. *IEEE Communications Surveys.*

[6] Yang G, Zheng J, Shi JH, Chen H. [2009] Energy Balance Hierarchical Data Aggregation Mechanism for Wireless Sensor Network. *WASE International Conference on Information Engineering*.

[7] Guiyi Wei, Yun Ling, Binfeng Guo, Bin Xiao, Athanasios V. [2011] Vasilakos Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter. *Computer communication* 34: 793–802.

[8] Hyuntea Kim, Jaebok Park, Giwhan Cho. [2007] Statistical Data Aggregation Protocol based on Data Correlation in Wireless Sensor Networks. *International Symposium on Information Technology Convergence*.

[9] Li G, Wang Y. [2011] An Efficient Data Aggregation Scheme Leveraging Time Series Prediction in Wireless Sensor Networks. *International Journal of Machine Learning and Computing*, www.ijmlc.org/papers/55-A836.pdf

[10] Rajesh G, Vinayagasundram B, Saravana Moorthy G. [2014] Data Fusion in wireless sensor network using Simpson's 3/8 rule, *International Conference on Recent Trends in Information Technology*.

[11] Liu L, Yang H, Lai M. [2005] Electricity price forecasting model based on chaos theory. *Power Engineering Conference (IPEC), The 7th International Conference on*,

[12] Liu H, Huang D, Wang Y. [2011] Chaotic Dynamics Analysis and Forecast of Stock Time Series. *International Symposium on Computer Science and Society* (ISCCS),

[13] Ci S, Sharif H. [2012] Performance comparison of kalman filter based approaches for energy efficiency in wireless sensor networks, *In proceedings of the 3rd ACS/IEEE International Conference on computer systems and Applications* (AICCSA'05), *IEEE, Cairo, Egypt, 58–65.*

[14] Giona M, Cimagalli V, Morgavi G, Perroiie A. [1990] Local prediction of chaotic time series. *Proceedings of the 33rd Midwest Symposium on Circuits and Systems.*

[15] http://db.csail.mit.edu/labdata/labdata.html.

[16] http://www.abarim-publications.com/ChaosTheoryIntroduction.html#.VS94WdyUeNM

[17] http://en.wikipedia.org/wiki/Chaos_theory