REDR: **R**eliable and **E**fficient **D**ata **R**eduction Techniques for Single and Multi-Modal WSNs

Mohamed O. Abdel-Aal
Demonstrator, Department of Communication and Electronics, Faculty of Engineering, Port Said University
Mail: Mohmamedyd@yahoo.com
Tel: +20122622256

Rabie A. Ramadan
Assistant Professor, Department of Computer Science, Faculty of Engineering, Cairo University
Mail: rabie@rabieramadan.org
Tel: +20167643227

Ahmed A. Shaaban
Assistant Professor, Department of Communication and Electronics, Faculty ofEngineering, Port Said University
Mail:dessouki2000@yahoo.com
Tel: +20101083478

Mohamed Z. Abdel-Meguid
Professor, Department of Systems and Computer Science, Faculty of Engineering, Al-Azhar University
Mail: azhar@mailer.eun.eg
Tel: +20102890933

Abstract—Wireless Sensor Networks (WSNs) suffer from many limitations such as the computing capabilities, and the allocated bandwidth. However, the limited energy source is the dominant factor where energy starvation occurs due to the large number of messages need to be transferred through the network. Several techniques for minimizing the data sent by different nodes to the base station are proposed in order to maximizing the WSN lifetime. Our contributions in this paper are: 1) increasing the lifetime of the wireless sensor nodes, 2) taking care of sensors reliability in which nodes reports accurate data, 3) introducing predication techniques for data reduction in WSN including Exponential Smoothing Predictors (ESPs) (single and double exponential algorithms) , 4) introducing a fuzzy logic approach for data reduction, 5) applying our data reduction proposal for a real time application such as greenhouse monitoring, 6) the proposed approach is applied for both single and multi-modal WSN. The proposed solution for Reliable and Efficient Data Reduction (REDR) in WSN is pronounced as “red r”. Through large number of experiments, our approach is tested through real data and WSN prototype as well as through simulation. Additionally, we demonstrate the importance of Wireless Sensor Networks (WSNs) in Greenhouses Applications by using single WSN prototype for different applications. Moreover, we make use of a simple Fuzzy Logic Controller (FLC) in determining the probability of plants’ infections by certain diseases in order to make the right decisions.

Keywords—Wireless Sensor Network, Data Reduction technique, Reliability, Multimodal WSN, Fuzzy Logic, WSN Prototype, and Exponential Smoothing

I. Introduction

The need for a new technology model that collects and analyzes data from the surrounding environment has led to the emergence of Wireless Sensor Networks. WSN is a network of wireless battery-powered sensors that monitor our environment, our machines, and even us. These sensors are self-contained units equipped with a radio transceiver, a small processor, and an energy source. These nodes are responsible for measuring a certain quantity, creating packets, and sending these packets to a base station via multi hop transmission. However, it was found that the radio communication among the different nodes consumes the majority of the available energy. For instance, Verdone [1] found that the energy cost of transmitting 1 Kb of information a distance of 100 m is approximately the same as that of executing 3 million instructions [2]. Consequently, the radio messages are the main reason for shortening the node's lifetime. Many devices such as Mica2 and MicaZ that are used in WSN run on two AA batteries. Depending on the activity level of a node, its lifetime may run onlyfor a few days if no power management schemes are used. This paper focuses on the sensed data not on the overhead data reduction since it has the major effect on the lifetime of WSN. We consider two different types of WSNs which are single and multi-modal networks. In single modal networks, each sensor is assumed to measure a single feature from the sensed environment while in multimodal WSNs, a sensor may sense multiple features at the same time such as temperature, humidity, and pressure.

Since each node has limited battery power, efficient power consumption is a challenging problem in WSN.During signal propagation, the signal decays as $r^{-\alpha}$ with transmission range r, where α is the loss exponent of the signal[2].The limited power and signal loss during propagation impose fundamental constraints onthe operational lifetime of the WSN. In many applications, it is expected that each sensor node last for a long time because in most of the cases WSNs are usually used in remote areas ; recharging and replacing power supply units is difficult.For this reason, energy efficient models have to be employed to reduce the wasteful power that is consumed in the radio communication. Therefore, we have to find an optimal solution that maximizes the lifetime of the whole network while reducing unnecessary power consumption. One way is to reduce the number of messages to be sent by sensors to the sink node with guaranteeing some data accuracy. This can mathematically formulated as follows:

$$min \sum_{t=0}^{L} \sum_{s=1}^{n} M_{st} X_{st} \forall_{s,t} \quad (1)$$

Provided that:

$$L_{st} \geq \mu \forall_s \quad (2)$$

$$|T_{st} - T_{s(t-1)}| \leq \gamma \forall_s \quad (3)$$

Equation (1) tries to minimize the number of messages M_{st} that can be sent by a sensor s starting from $s=1$to n at any time t. At the same time, a sensor s is considered died when its lifetime or residual energy L_{st} is reached a certain threshold μ as stated in (2). $X_{st}$is a binary variable that is set to *1* when a message is to be sent and *0* otherwise. On the other hand, it is necessary to maintain the accuracy of the data T_{st} sent by a sensor s at time t high and suitable for the application's requirements as described in (3).

However, the data reliability, in this case, depends on the threshold value defined by the WSN user (γ *in our case*). The term reliability in this context means the accuracy of the data sent to the sink node. Such data has to be as much as possible the same as the sent by the sensors. The sink node will be working on values other than the original readings of the sensor nodes. Therefore, it is obvious the reliability and data reduction are contradicting terms since the reliability requires sending every piece of information to the sink node; while the data reduction and energy conservation require the minimization of the number of messages that are received by the sink node for other nodes. Our proposal in this paper considers both the data reliability as well as the data reduction for the purpose of maximizing the overall network lifetime.

In this paper, exponential smoothing techniques are used as predictors to predict the next sensors readings. This type of predictor is used in many applications such as Inventory Control Applications, Tracking, Finance,and many other applications. Also, fuzzy logic technique is used in the multi-modal WSN to produce a certain decision at which the node has to send its information to the base station for evaluation. Also, we investigate the data reliability of the predicted values.Throughout this paper, we experiment the performance of the proposed algorithms based on two categories of experiments. The first category of experiments utilizes a simulation to WSN based on different network topologies as well as communication ranges and sensing ranges. Such experiments simulate outdoor WSN suitable for critical applications such as battle field and habitat monitoring. Secondly, we test the exponential smoothing predictors on a real WSN where National Instruments (NI) WSN kit is used to examine the effectiveness of our approach.

Additionally, we propose two schemes for monitoring system inside greenhouses. In the first one, we add an additional mission for the deployed sensor nodes by implementing a fuzzy logic controller in these nodes to determine either the probability of infection (PoI) for each disease. However, the nodes will not report any information to the sink node as long as the output of the fuzzy logic algorithm lies below a certain safe threshold. Consequently the nodes' lifetime maximization is achieved through a threshold based approach rather than employing the REDR algorithm. In the second approach, sensor nodes are deployed inside the greenhouse where the function of these nodes is to report the environmental parameters without any further processing except the REDR algorithm that runs for maximizing the lifetime of the deployed sensor nodes. Simultaneously, the sink node is connected to a host controller which processes these measurements using fuzzy logic controller to determine the probability of infection for the diseases mentioned below in Table 1. Afterwards, the framers will make a decision based on the resultant probabilities. FLC is chosen because it consist several advantages compared to the other classical controller such as simplicity of control, low cost and the possibility to design without knowing the exact mathematical model of the process.

TABLE 1

Optimum conditions for some Ornamental diseases

Disease	Temperature	Humidity	Soil moisture	Nitrogen	Shadowing
Gray leaf Spot	70°F : to 95°F	Greater than 70%	Low	High	Semi-shadow
Botrytis leaf blight	70°F :77°F	Greater than 85%	High	High	Shadow
Downy Mildew	60°F : 74°F	Greater than 85%	High	Low	Shadow
Rust	59°F : 77°F	Greater than 55%	Low	Low	Sunlight

The paper is organized as follow: the following section elaborates on the related work; section 3 explains the main idea behind time series prediction techniques and fuzzy logic system used in this paper; section 4 shows the details of our experiments; section 5 describes the application of WSNs in detecting the ornamental plants diseases inside greenhouses.Finally, the paper concludes in section 6.

II. Related Work

Many research papers and surveys investigate different approaches that can be used to prolong the WSN lifetime either from the sensor node point of view or from the whole network point of view. In the first approach, lifetime was maximized by reducing the power consumed in the different operations inside the sensor node through using the dynamic voltage scaling, dynamic frequency scaling, asynchronous processors, ultra wideband for radio communication, and CMOS low voltage and low power wireless Integrated Circuit (IC). Others benefit from the surrounding environment by converting the ambient energy such as solar, vibration, and wind energies into electricity to power the sensor nodes as described in [3].

In the second approach, Hnin Yu, et al [4], for instance, enumerate four techniques for saving energy. The first one is the use of sensors' scheduling by which sensors alternate between sleeping and waking; the waking sensors sense events in their environments and the sleeping sensors avoid idle listening and overhearing. The second one is to aggregateseveral events into a single event at intermediate nodes to reduce transmissions.Network coding is the third solution in which the collected data are mixed at intermediate node and then encoding packets are sent instead of sending individual packets; consequently it reduces the traffic among the different nodes. In the last one, data collision are avoided to reduce the retransmission of packets; this can be achieved byscheduling nodes onto different sub-channels that are divided either by time, frequency or orthogonal codes through implementing communication protocols such as Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA) and Code Division Multiple Access (CDMA), respectively.

Energy conservation techniques ,such as the ones stated in [14][2], in general, are classified by Giuseppe et al [5] into duty cycle, data driven, and mobility-based techniques. Further, the authors classified the data driven techniques into in-network processing, data compression, and data prediction techniques as shown in Fig.1. As can be seen in the Figure, there are three types of data prediction algorithms which are stochastic, time series forecasting, and algorithmic approaches.

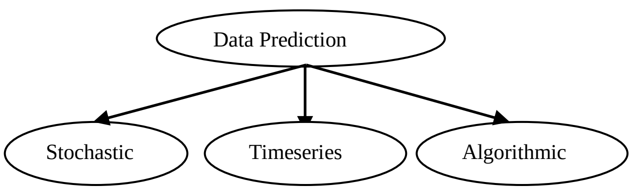

Fig.1. Classification of data prediction algorithms

Stochastic algorithms are based on probabilistic models such as correlation [6], Kalman filter [7] and Dynamic Probabilistic Model (DPM)[8]. Some of the time series algorithms are also utilized in the literature. For instance, Santini, et al.[9] proposed data reduction technique that uses Least Mean Square (LMS) adaptive algorithm. The LMS is an adaptive algorithm with very low computational overhead and memory footprint that provides excellent performance. It also does not require a priori knowledge or modeling of the statistical properties of the observed signals. Nicholas Paul, et al [10] implemented this algorithm on FPGA kit where they showed that this method managed to increase the network lifetime by almost 18,962.5 % when compared to an always on solution. However, sensors' data usually have a trend and might seasonal information that we can benefit from. Moving Average (MA) and Auto-Regressive prediction algorithms are also used in WSN. The authors use these prediction algorithms on each node to send only one value which is the predicted value after the nodes' buffer is full. Our approach is similar to these time series forecasting algorithms. However, we use the same model on the sensor nodes as well as on the sink node assuming that the sink node is more powerful to run the model for each sensor in the network. There are some other algorithmic or heuristic techniques for data prediction such as PREMON and buddy protocols.

III. RELIABLE AND EFFICIENT DATA REDUCTION TECHNIQUES

In this section, we present our approach, REDR, for reducing the number of reports between the different nodes. There are many prediction schemes such as Adaptive filters based predictors, Markov predictors, Linear Extrapolation, Neural network based predictors, and Exponential smoothing predictors. However, we investigate the data reduction technique based on the exponential smoothing algorithms since they are more appropriate if the data include rapid fluctuation or more complex patterns of behavior in addition, they need very small set of data to predict the incoming values, so they require small memory footprint. In addition they provide low computation overhead. Moreover, we investigate another computational intelligence technique by utilizing the Fuzzy Logic in data reduction.

A.Exponential Smoothing

Exponential smoothing is a mathematical-statistical method of prediction used in industrial engineering where the prediction is based on detecting significant changes in data. In exponential smoothing older data is given progressively-less relative weight whereas newer data is given progressively-greater weight.The reason for this is that the future may be more dependent upon the recent past than on the distant past. The method is effective when there is random demand and no seasonal fluctuations in the data. It is commonly applied to financial market and economic data, but it can be used with any discrete set of repeated measurements.

Data reduction is achieved by implementing identical predictors on sensor nodes as well as on the sink node. In a clustered network, the same predictor will be running on the cluster heads as well. Node's predictor is used to predict the sensor's readings, where the predictor on the sink node is initialized after the sensor node sends a number of readings to it. Therefore, on the sensor, we have two values, the sensor's reading and the predicted value. If the difference between these two values, which is called the error signal, is lower than a certain threshold value then the transmission between the two nodes is canceled, and the sink node stores the predicted value generated from its own predictor as the sensor reading at this time. On the other hand, if the error signal is greater than this threshold, then the sensor reading is transmitted to the sink node, and the sink node uses this value to update its predictor. For accurate prediction to the next readings on both sink and a sensor node, a sensor node runs exactly the same agent running on the sink node. The only difference that the sensor node agent does is the comparison between the estimated value and actual value to take the sending decision. In other words, although a sensor reads the actual sensing information but the sensor would not predict the next value based on the actual data, it acts exactly the same as the sink node. The, the next predicted value is compared to the actual (measured) value to compute the error difference between them. Once the error is greater than a certain threshold, the sensor node sends its actual reading to the sink node. Now, the sink as well as the sensor node runs their predications based on the new (actual) data.

There are three models of the exponential smoothing method based on nature of the data stream under consideration. We will briefly describe the three exponential smoothing models.

1) Single Exponential Smoothing (SES)

Robert G. Brown [11] proposed this idea in 1944 while he was working for the US Navy as an Operations Research analyst.This method is used when the data has a *mean*that is either stationary or changes slowly with time. In other words, for higher prediction accuracy SES is used when the data does not include trend or seasonality components. The simplest form of exponential smoothing is given by the formulae given in equation (4) [20].

$$F_{t+1} = \alpha y_t + (1-\alpha) F_t \qquad (4)$$

Where $F_{t+1}$is the prediction for the next period, α is the smoothing constant, y_t is the measured value in period t, and F_t is the old forecast for period t. If we recursively apply the smoothing equation to F_{t+1}, we get:

$$F_{t+1} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \qquad (5)$$
$$.... + (1-\alpha)^{n+1} F_{t-n}$$

as time passes the smoothed statistic F_t becomes the weighted average of a greater and greater number of the past observations y_{t-n}, and the weights assigned to previous observations are in general proportional to the terms of the geometric progression $\{1, (1-\alpha), (1-\alpha)^2, (1-\alpha)^3, \ldots\}$.

The smoothing constant must satisfy the following inequality $0 < \alpha < 1$, if α is chosen close to 1, it willhave less of a smoothing effect and give greater weight to recentchanges in the data, while values of α closer to zero have agreater smoothing effect and are less responsive to recent changes.For higher accuracy, the value of α which satisfies the smallest Mean Squared Error (MSE) is chosen for use in producing the future predictions. The initial value of F_t plays an important role in computing all the subsequent values. One option is to assign the first observation as an initial value. However, we can get more accurate initialization by averaging the first five observations.

2) Double Exponential Smoothing (DES)

During the 1950s, Charles C. Holt[11] developed a method for exponential smoothing of additive trends. the predictor's operation is initiated by decomposing the data under consideration into level and trend signals afterwards these signals are smoothed using (6) and (7) while the predicted value are found using (8).

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} - b_{t-1}) \qquad (6)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1} \qquad (7)$$

$$F_t = L_t + m b_t \qquad (8)$$

Where L_t is the estimate of the level at time t, y_t is the actual value of series in period t, α is the data smoothing factor, $0 < \alpha < 1$, β is the trend smoothing factor, $0 < \beta < 1$, b_t is the estimate of the slope of the series at time t, and m = periods to be predict into the future. In(6) L_t is adjusted directly for the trend of the previous period, b_{t-1}, by adding it to the last smoothed value, L_{t-1}. This helps to eliminate the lag and brings L_t to the appropriate base of the current value.Afterwards,equation (7) evaluates the trend at time t, which is expressed as the difference between the last two values. The equation is similar to (4), but here applied to the updating of the trend.The smoothing constants α and β are chosen independently between 0 and 1, where their values must satisfy the smaller MSE. To start the process, both L_1 and b_1 must be initialized by setting $L_1 = y_1$ and $b_1 = y_2 - y_1$. Thus, no forecasts can be made until y_1 and y_2 have been observed. By convention, we let $F_1 = y_1$.

3) Triple Exponential Smoothing (TES)

Holt developed a method for smoothing seasonal data in 1960 and Winters tested his method with empirical data. They became known as the Holt–Winters forecasting system. We will not state the mathematical model since it is only suitable for data stream with trend and seasonality. We limit the research in this paper to the previous two prediction models since the real data that we have has no seasonal features. However, we plan to study the Holt-Winter's method when we get access to sensors' seasonal data.

The work in this paper investigates the lifetime of the sensor network in three modes of operations:

a. Normal mode where no power saving technique is used in the WSN. We call this method a naive method.
b. Single modal in which the sensors will collect only one environmental phenomenon such as temperature or humidity. This model will be tested twice one of with SES while the other using DES.
c. Multimodal WSN where the sensors will collect multiple phenomena such as temperature, humidity and pressure. There are two techniques to send thesemeasurements from a sensor node to a sink node, either using one data message for all measurements or using multiple messages sending each phenomenon in a separate message. Each case will be tested along with SES and DES as well.

B. Fuzzy Logic

Lotfi A. Zadeh proposed the concept of fuzzy logic in 1965 [12]. Fuzzy logic is a multi-valued logic, where it formalizes reasoning when dealing with vague terms. The decisions are not limited to either true or false, or as with Boolean logic either *0* or *1*. Therefore,fuzzy logic algorithms take into consideration thedegrees of truthfulness and falsehoods. For instance,the Boolean logic is not only *0* and *1* but also all the numbers that fall in between.

The process of fuzzy logic is shown in Fig.2where crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Finally, the resulting fuzzy output is mapped to a crisp output using the membership functions in the defuzzification step as shown in Fig.2.

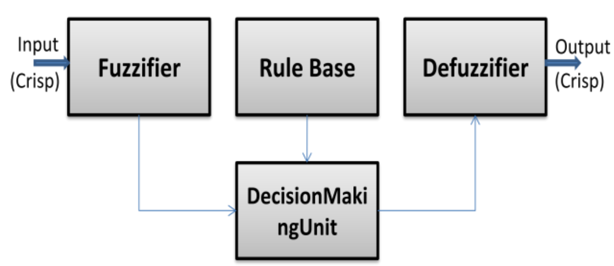

Fig.2.A Fuzzy logic System [13].

Linguistic variables are a representation for the system's input or output variables where they convert the numerical values into words or sentences from a natural language. While the membership functions are used in the fuzzification and defuzzification steps of a fuzzy logic, to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. Therefore, sensors will send only the decision not the actual data which could be as simple as a single bit.

In the multi-modal WSN, multiple environmental phenomena have to be collected and sent to the end user for evaluation; however, this will accelerate the depletion of the sensor node's energy. Fuzzy logic is a proposed solution for this problem through minimizing the transmissions in the network. Hence, adding intelligence to the nodes will convert their function from sending measurements to sending information that is generated based on the measured parameters. In this paper, the input to the fuzzy logic algorithm is values of temperature, humidity, and light intensity at a certain area. These values will be transformed to linguistic variables in order to be processed using the fuzzy rules. The output of the fuzzy logic algorithm is a value that represents the probability of certain event to occur. Therefore, the sensor nodes will not turn on its transceiver unless this probability exceeds a threshold set by the user.

IV. EXPERIMENTS AND SIMULATION RESULTS

In this section, we test the performance of REDR for maximizing the WSN lifetime through simulations and practical experiments. Several tests are made on the WSN before and after the implementations of exponential smoothing predictors, fuzzy logic algorithm, and other simple algorithms based on comparing present and past measurements. This section is divided into two main sub-sections. In the first one, we will present the simulations of WSN when applying our approaches. While in the second one, we will show the results of the practical experiments.

A.Simulation Results

A WSN simulator is developed to investigate how REDR reduces the transmitted data among the sensors as well as between the sensors and the sink node. We modified the simulator developed by David J. Stein [14]. The simulator is written in C# and used to detect and report certain events across certain areas. The author simulated the WSN as a connected graph and sensors always transfer their data towards the sink node. Sensors decisions in sending and/or forwarding messages are based on their residual energy as proposed in [16].The simulator is modified to cope with our desired mission, where the event is changed from detecting moving objects to sensing environmental phenomena such as temperature, humidity, or pressure.

For the purpose of accuracy, we implemented a similar energy model to the one used in [17]. The energy model used to compute the consumed energy during the sending and receiving by each node. The first order radio model stated that the energy for transmitting 1 bit data over distance d is:

$$E_t = a_1 + a_2 * d^k \qquad (9)$$

where a_1 is the energy spent by transmitter electronics (a_1 = 50 nJ/bit), a_2 is the transmitting amplifier (a_2 = 100 pJ/bit/m^2), and k is the propagation loss exponent. Therefore, sending a 1 bit is computed using equation (9). On the other hand, the energy for receiving 1 bit data E_r is equal to the energy spent by receiver electronics and it is usually considered as a_3 = 50 nJ/bit. This model is used throughout all of our experiments.

Our simulator runs on a Windows7 machine with 2.13 GHz processor and 2 GB RAM. 35 nodes are randomly deployed in an area of 460 X 300 cm. It is assumed that nodes are equipped with a 1 cm^3 of non-rechargeable lithium battery (at maximum energy density of 2880 J/cm3 or 800 watt hour per liter) were to consume 100 µW of power on average. The lifetime of the WSN is considered as the period between the start of simulation process till the depletion of at least one sensor node from its energy. In our experiments, we believe that it will not be practically to consider the network lifetime as the running time of the simulator until a node's energy is depleted. Thus, we count the lifetime based on the number of iterations that the simulator takes till a node dies. The iteration is described as a complete pass over all the nodes having data to be sent and/or forwarded till their data reaches the sink node. In the following subsections, we will test the influence of the data reduction technique based on exponential smoothing predictors, fuzzy logic algorithm, and threshold and tolerance based algorithms on the lifetime of WSN using the described simulator.

1) Data Reduction Technique Based On Exponential Smoothing Predictors

We test our data reduction proposal based on single and multiple smoothing predictors for both single modal and multimodal WSN. But at the beginning, we have to examine the predictors separately to determine the best smoothing constant found in (4), (6), and (7). Therefore, our experiments are categorized into two categories which are experimental category based on real data and simulation results based on simulation to the WSNs environment.Throughout the next sections, we will elaborate more on these categories.

a) Predictors' parameters and reliability test

In the first set of experiments, the exponential smoothing predictors were tested on a set of real world data which is publicly available at [18]. Every 31 seconds, humidity, temperature, light intensity and voltage values were collected from 54 Mica2Dot sensor [18] nodes that were deployed at the Intel Berkley Research Lab between February 28th and April 5th, 2004. These measurements are used as input to SES and DES algorithms in order to examine many values for the coefficient constants and find the best values that minimize MSE. The results given in Figs.3 and 4, show that the larger the smoothing constant the lower the Mean Square Error (MSE). In addition, we noticed that the MSE in case of temperature values is smaller than that in case of humidity.

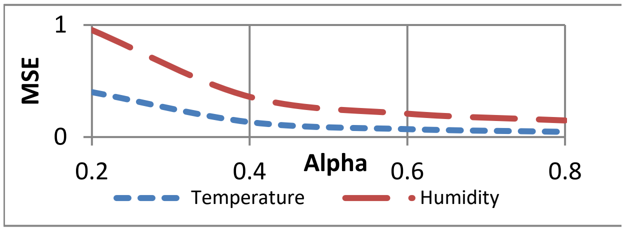

Fig.3.MSE decreases as α increase.

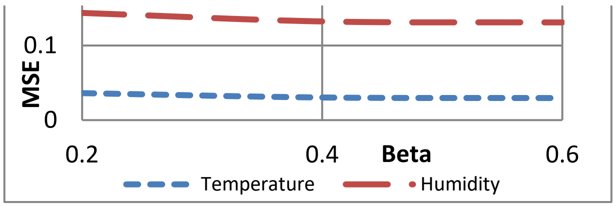

Fig.4.MSE decreases as β increase.

As can be seen in Figs.3 and 4, starting from β=0.4 the MSE values are almost the same. Therefore, we selected β to be 0.4 and α to be 0.6. The reason behind choosing α to be 0.6 not 0.8, as given in Fig.4, based on our observation in DES, MSE is approximately constantbeyond α=0.6. Therefore, to standardize our experiments, we set α to 0.6.Also, the MSE in case of temperature values is smaller than that in case of humidity based on the used data. Therefore, exponential smoothing predictors are more accurate if the data under consideration represents temperature values rather than humidity values.

The reliability of the measurements and the predictions, from the accuracy point of view, is estimated based onthe Standard Deviation (SD). It is sometimes evaluated as a percent of the mean, in which case it's known as a Coefficient of Variation (CV). When the measurements of a single subject are repeated to determine the mean and SD, the resulting coefficient of variation is an important measure of reliability. As shown below in Fig. 5, CV for the humidity readings are inversely proportion with the smoothing constant (Beta) in case of DES. However, the readings are linearly proportion with the smoothing constant (Alpha) in case of SES. So we can conclude that the lower the smoothing constant the higher the data reliability if the SES is used, while the higher the smoothing constant the higher the data reliability if the DES is used.

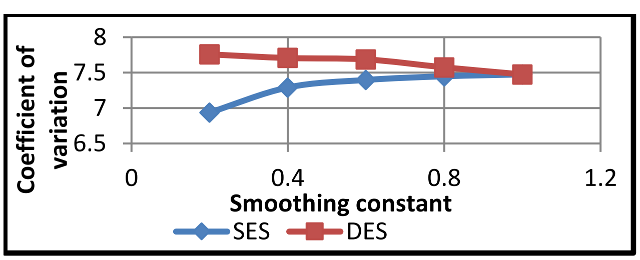

Fig.5.Coefficient of variation in case of humidity readings

b) Simulation results

In this subsection, two case studies for simulating the WSN are considered. In the first one, we investigate the lifetime of single modal WSN in which temperature or humidity is measured and sent to the base station. While the second one determine the lifetime of multimodal WSN in whichtemperature and humidity and light intensity are collected and sent to the base station. The smoothing constants are set to α= 0.6, β = 0.4 according to the results obtained from previous sections. In addition, the threshold varies from 0.0 to 2, so the sensors will not send its readings as long as the error signal is smaller than the given threshold.

The results presented in this section is the average values over 1000 runs to the simulator with different settings including different network topologies. WSNs used in our simulator are considered to be flat where no clustering techniques are used. In addition, the measurements are simulated based on the real measurements extracted from [18]. Also, since our approach is over performing the naive model with 1000's of times, we will use the logarithmic scale for comparisons.

1. Single modal WSN

Here, we compare our reduction approaches to the naïve communication method between the sensors and the sink node where each sensor sends its information or forward others information in a separate message. Fig.6 depicts the comparison among three different algorithms which are the naïve, SES, and DES. The three algorithms are compared in terms of their lifetimes with different threshold values. As can be seen in the figure, the SES over performsDES for low threshold values. However, they produce almost the same performance for high threshold values. At the same time, both of them over perform thenaïve modal for all threshold values.This means that predictors' error is small relative to the threshold value. In addition, with increasing the threshold value, the lifetime of the network increases with large percentage since there is no need for sensors to flood the network with messages. However, the reliability of the data will be degraded.

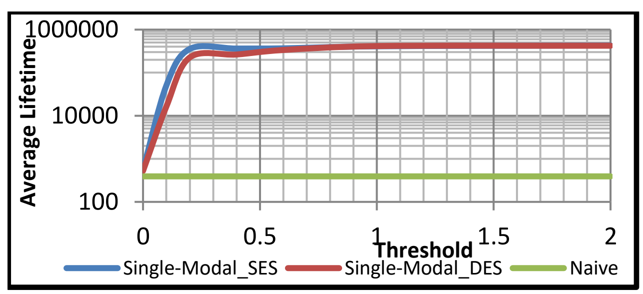

Fig.6. Lifetimes in Single Modal WSN (Logarithmic scale)

2. Multi-modal WSN

In this set of experiments,multimodal sensors have been used where a sensor may report more than one feature from the monitored field. We test our approaches with two different settings which are single packet and multi packet; in single-packet settings, the monitored data are collected and sent all together in one packet to the sink/cluster head node. While in multimodal multi-packets, each measured phenomenon is sent within a separate packet to the collector node.

For single packet settings shown in Fig.7, SES seems to perform the same as DES. At the same time, their performances tend to be much better than the naive model. In fact, the lifetime increases almost linearly when predictors' are used until threshold reaches 0.3 then it becomes fairly constant beyond this threshold.

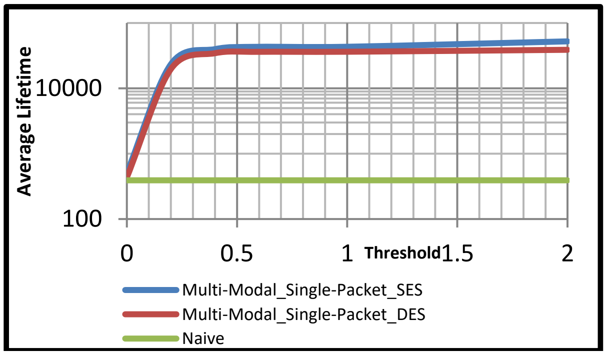

Fig. 7.Lifetimes in Multi-Modal WSN using real sensor readings (Single-packet settings).

While in the second setting as shown in Fig.8, a multimodal WSN is simulated with sending the sensed features in separate messages when it is needed. The performance of our approach is similar to the one presented in Fig.7 except the fact that SES over performs DES for threshold values below 0.4. However, we need to compare the performance of REDR using Exponential smoothing predictors and that using Adaptive filters described by Santini, et al. in [10]. As shown in Fig. 9, REDR technique using ESPs over performs REDR using LMS adaptive filter.

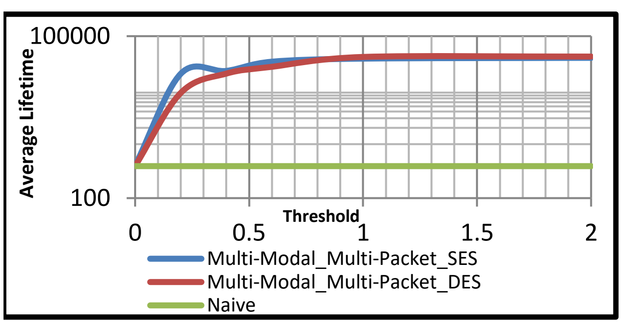

Fig.8. Lifetimes in Multi-Modal WSN using real sensor readings (Multi-packet settings)

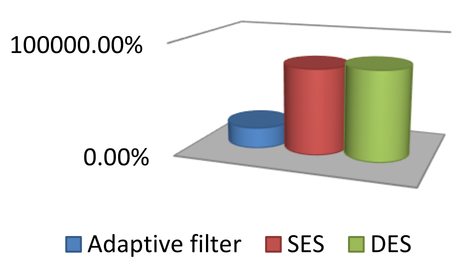

Fig. 9 Comparison between REDR using ESP and REDR using LMS adaptive filters

2) Data Reduction Based on Fuzzy Logic Algorithm

Fuzzy logic is used as an alternative solution to the problem of power consumption in WSNs through minimizing the number of messages sent to the end user. In this case, sensor nodes turn off their transceivers until a real danger is probably exist. This danger is sensed based on temperature, humidity and light intensity measurements.

As shown in Fig. 10, "Low", "Medium" and "High" terms are used as the linguistic input variables. In the same manner, "Very Low", "Low", "Medium", "High" and "Very High" terms are used for the linguistic output variable (probability of fire). On the other hand, the fuzzy rules are set empirically. For instance, table 2 shows sample of the fuzzy rules used in the defuzzification process to determine the probability of fire.

In addition,the center of gravity for singletons method is used to determine the output (probability of fire occurrence).

TABLE 2
SAMPLE FUZZY RULES

Temperature	Humidity	Light Intensity	Output
L	H	L	VL
L	M	H	L
H	M	M	H
H	L	H	VH
:	:	:	:

The simulator parameters are the same as those in the previous section, and the real data extracted from Intel Berkley Research Lab are used as the measurements of the sensor nodes.

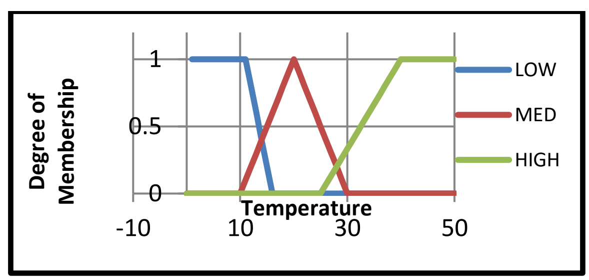

Fig. 10. The membership function for the input variable

As shown in Fig.11, the average lifetime of the WSN increases as the probability of fire increase except for the region between 55% and 59%, the lifetime is approximately constant. Through comparing Figs.6, 7, and 8 with Fig.11 and from Fig. 12, it is obvious that, the average lifetime of the WSN based on fuzzy logic system at 60% probability of fire, for instance, is fairly double that of the single modal WSN and is approximately 17 times that of the multi modal WSN based on exponential smoothing predictors.

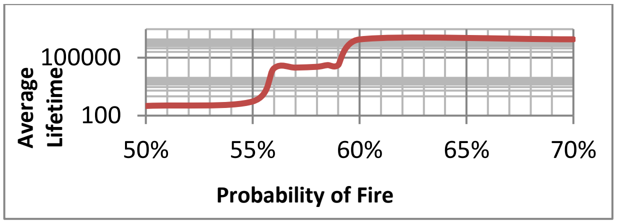

Fig. 11. The average lifetime of WSN for different probability of fire

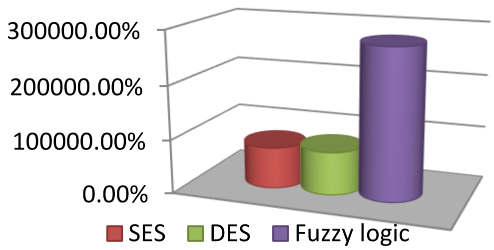

Fig. 12. comparison between REDR using ESPs and Fuzzy logic algorithm

B. Experimental Results

In this section, we test the exponential smoothing predictors on a real WSN. National Instruments (NI) WSN kit is used for this purpose [18], where its sensor nodes are programmable so that we can download our algorithm on its flash memory and then watch its behavior for a period of time. The kit contains an NI WSN-9791 Ethernet gateway, two battery powered programmable nodes (NI WSN-3202 ±10 V analog input node, and NI WSN-3212 thermocouple input node). The gateway works as a bridge between the IEEE 802.15.4 wireless network and the wired Ethernet network. The default behavior of an NI node is to sample all channels and transmit every sample acquired to the gateway. However, we added intelligence for the nodes to increasetheir lifetime by extending the transmit interval which is equal to the time period between the successive transmissions of a certain sensor node.

The measurement nodes are programmed using NI LabVIEW which is graphical programming environment used to develop measurement, test, and control systems using graphical codes. These codes are drawn insideVI block diagrams whichinclude data processing modules, and predefined mathematical models such as adders, substructures, integrators, etc. used in a mixed modality platform.Fig.13depicts the Virtual Instrument (VI) block diagram for DPS predictor. We downloaded it to the sensor nodes but at the same time an identical predictor is running on the host controller. The lifetime of the node is determined based on the sample interval, and the transmit interval. In our experiment, we set the sampling rate to constant value, for instance 5Hz, while the transmit interval is adaptive where this period is determined based on the error between the actual measurements and the predicted values. For example, if the error signal is below 0.5 degree, the transmit interval is set to 5000 second. While if the error signal is below 1 degree, the transmit interval is reduced to 50 second. However, if the error increased beyond 1 degree, the nodes transmit every 1 second. But from our observations, the nodes transmit the measurement every 5000 second due to the ability of the predictor to follow the real values especially in case of temperature measurements whose values change slowly. From the LabVIEW WSN pioneer performance benchmark, it is stated that the lifetime of the nodes is 5.3 month when the transmit interval is set to 5 seconds. On the other hand, our approach increases the lifetime to 27 month for transmission interval of 5000 second. However, we can obtain longer lifetime if we extend the transmit interval beyond 5000 second.

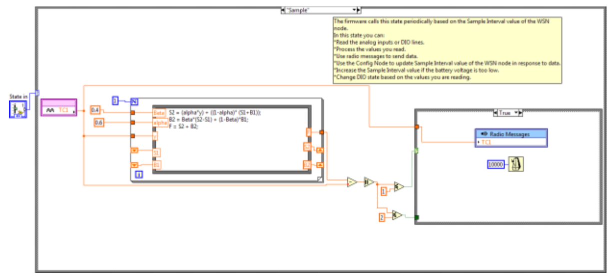

Fig. 13. Controlling the transmit time by the DES predictor (VI)

V. Real Life Application: Greenhouses

In this section, we describe our WSN architecture for Greenhouse monitoring. Two schemes, calledGreenSense_1 and GreenSense_2, are proposed to detect and report the diseases that may infect the ornamental plants inside Greenhouses [20].

A. GreenSense_1

It simply uses the deployed WSN inside greenhouse to compute thePoI of the ornamental plants inside greenhouse. Therefore, Fuzzy Logic Controller (FLC) will be implemented on each node to estimate these probabilities. However, a threshold will be defined to prevent the radio communication for low values of PoI. To implement these controllers, we decomposed the FLC into five controllers. Each one is responsible for estimating PoI by a certain disease. Mamdani reasoning technique will be used for the inference system since it is more intuitive, has widespread acceptance, and is better suited for human input. In addition, the rules are created based on the optimum environmental parameters that cause the infection by ornamental plants diseases. Table 3 depicts sample of these fuzzy rules for Gray leaf Spot.

These controllers will be implemented on each sensor node inside the greenhouse. Fig. 14 depicts the Virtual Instrument (VI) block diagram. For simplicity, we used only three linguistic variables for the input values “Low”, “Medium” and “High”. Afterwards, the output variable is determined based on the temperature, humidity, nitrogen, and soil moisture. In this code, the node will not transmit any radio messages until the probability of disease occurrence increase beyond 0.6.

TABLE 3
SAMPLE FUZZY RULES

Temperature	Humidity	Soil moisture	Nitrogen	Light intensity	PoI
Cool	Low	high	Medium	Semi-shadow	Low
Cool	Low	Medium	Medium	Sunlight	Medium
Warm	High	Medium	High	Shadow	High
:	:	:	:	:	:

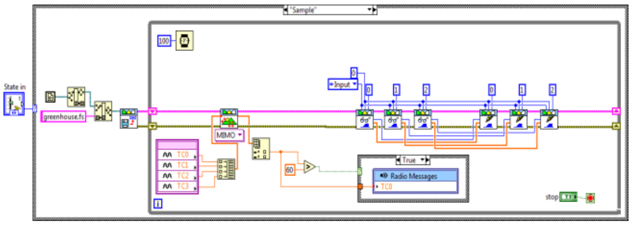

Fig. 14.Block diagram of the implementation of SES in LabVIEW (VI)

B. GreenSense_2

In this scheme, sensor nodes reports environmental parameters without any processing. However, FLC will be implemented on the controller connected to the sink. The FLC will be prototyped by using the Fuzzy Logic Toolbox in Matlab [22]. Through this toolbox, we can set up prototype triangular fuzzysets for the fuzzy variables. The fuzzy inference consists of 243 fuzzy rules. The rule base is viewed graphically as shown in Fig. 15 in which different inputs can be chosen to see what the output will be.

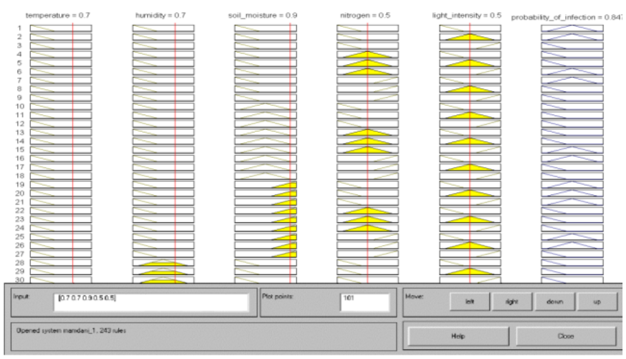

Fig. 15. The inferred output of the MISO system

Figs. 16 and 17 show examples for the control surface which is the output plotted against the two inputs, anddisplay the range of possible defuzzified values for all possible inputs. It is aninterpolation of the effects of the entire rules. As shown in Fig. 16, the probability of infection increases as the humidity and temperature increases. While in Fig. 17, the probability of infection is small for normal level of nitrogen. But this probability increases sharply when the level of nitrogen is very small or even very high.

The proposed FLC can be optimized through using the concept of Adaptive Neuro Fuzzy Inference System (ANFIS)[22]. The function of ANFIS is to allow the fuzzy system to learn by taking the FIS and tuning it with the back-propagation algorithm based on some collection of input-output data.

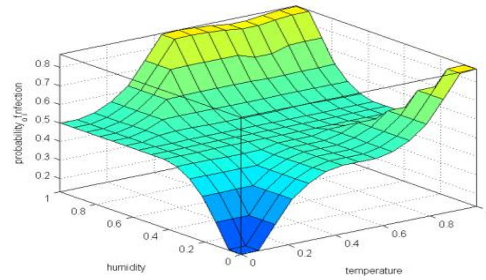

Fig. 16.Control surface for temperature and humidity

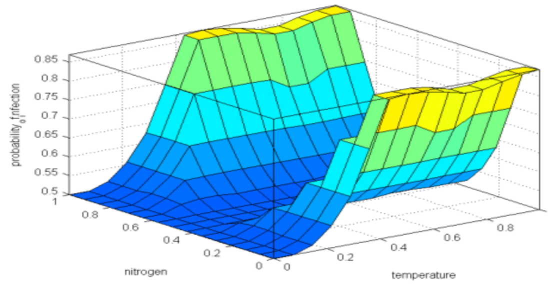

Fig. 17.Control surface for nitrogen and temperature

The main reason for employing ANFIS is the huge number of rules that will deteriorate the performance due to the computational overhead. ANFIS provide the option of generation a FIS with a smaller number of rules through subtractive clustering method. This method partitions the data into groups called clusters, and generates an FIS with the minimum number rules required to distinguish the fuzzy qualities associated with each of the clusters.As shown in Fig. 18the number of rules is reduced from 243 rules to only 10 rules.

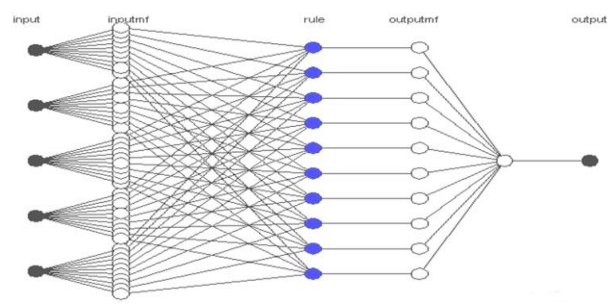

Fig. 18. ANFIS with Sub Cluster

GreenSense_2 is tested using a simple Simulink model as shown in Fig. 19. The same membership functions, fuzzy rules, and defuzzification method are utilized as those mentioned above. On the other hand, the inputs are simulated as proposed in [23]. For instance, temperature readings are simulated as a sine wave to present the day and night temperature changes. The amplitude of the wave is 5 degree plus a fixed bias of 30 degree with a frequency of 0.2618 rad/h. This frequency is measured according to a time period of 24 hours.

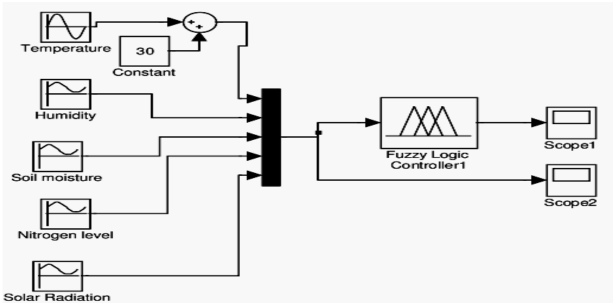

Fig. 19 Simulink model for fuzzy logic controller

As can be seen in Fig. 20, the probability of infection is a pattern that is repeated every day according to the stimulus. These probabilities are generated based on the fuzzy rules. In addition, these rules are constrained by the suitable environmental conditions for the infection by a Gray leaf Spot disease as an example for illustration.

Fig. 20 Probabilities of infection of Gray Spot Leaf

VI. CONCLUSION

In this paper, the impact of applying the data reductiontechniques on the lifetime of WSNs is investigated. The approaches utilized single and double exponential predictors, fuzzy logic algorithm, and threshold and tolerance e approaches to reduce the number of messages transmitted between the nodes and their neighbors as well as between the nodes and the sink node. Our approaches also targeted the reliability enhancement in sensors data. We tested the performance of our approaches compared to the naïve model where each sensor sends its sensed data in a separate message to the sink node.The simulation results showed that such approaches will increase the lifetime of the network up to 100s' of times consequently the in-network traffic is minimized. In addition, the practical experiment confirmed the results obtained from the simulation. Moreover, the results showed that the data reduction based on fuzzy logic algorithm provides the maximum savings in the nodes' power. In addition if data reduction based on ESPs will be used, it's recommended to used DES rather than SES since it achieves two important criteria including the maximizing the lifetime and improved data reliability. Finally, we proposed two prototypes for monitoring and detecting the diseases that infect the ornamental plants inside greenhouses.In the future work, we plan to design and fabricate our own sensor nodes that fit the improved features. In addition, we will investigate other approaches to maximize the WSN lifetime.

REFERENCES

[1] Anastasi,G., Conti,M., Di Francesco,M., Passarella,A. (2009). Energy Conservation in Wireless Sensor Networks: a Survey. Ad Hoc Networks, Vol. 7, No.3.

[2] Biljana Stojkoska, Dimitar Solev, and Danco Davcev , "Data Prediction in WSN using Variable Step Size LMS Algorithm," The Fifth International Conference on Sensor Technologies and Applications - SENSORCOMM 2011, 2011

[3] Bonnie Dorr, Kevin Knight, Daniel Marcu, Lori Levin, Alex Waibel , Kristin Procoda, Ace Sarich and Seiichi Yamamoto, (2003).10 emerging technologies that will change the world. Technology Review.Vol. 106, no.1, pp. 33-49.

[4] C. Leondes(1999).Fuzzy theory systems: techniques and applications. Vol. 1, Academic Press, Chapter 5.

[5] Chang,J.,and Tassiulas,L. (2004). Maximum Lifetime Routing In Wireless Sensor Network. IEEE/ACM Trans. on Networking, Vol. 12, no. 4, pp. 609-619.

[6] Chu,D., Deshpande,A., Hellerstein,J.M., andHong,W. (2006). Approximate Data Collection in Sensor Networks usingProbabilistic Models. In Proc. of 22nd International Conference on Data Engineering (ICDE06), p. 48, Atlanta,GA, April 3-8.

[7] Crossbow. Available at: http://www.xbow.com/, accessed Jan. 2011.

[8] Gardner,E. (2006). Exponential smoothing: The state of the art Part II. International Journal of Forecasting 22.

[9] Heinzelman,W., Chandrakasan,A., and Balakrishnan,H. (2000). Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In Proc. of the 33rd Hawaii International Conf. on System Sciences-Vol. 8, p.8020, Jan. 04-07.

[10] Intel Lab Data. http://db.lcs.mit.edu/labdata/labdata.html, accessed Jan. 2011.

[11] Jain, E.,Chang,Y., Wang, Y. (2004). Adaptive Stream Resource Management Using Kalman Filters. In Proc. of ACM International conferecne on Management of Data (SIGMOD2004), Paris (France), pp. 11-22, June 13-18.

[12] Javadi,P., Tabatabaee,A., Omid,M., Alimardani,R. and Naderloo,L. (2009). Intelligent Control Based Fuzzy Logic for Automation of Greenhouse Irrigation System and Evaluation in Relation to Conventional Systems.Publishedat the World Applied Sciences Journal 6: 16-23, ISSN 1818-4952.

[13] Kanagal, B. and Deshpande,A. (2008). Online Filtering, Smoothing and Probabilistic Modeling of Streaming Data. In Proc. of 24th International Conference on Data Engineering (ICDE 2008), Cancún, México.

[14] Msechu, E.J.; Giannakis, G.B.; , "Sensor-Centric Data Reduction for Estimation With WSNs via Censoring and Quantization," Signal Processing, IEEE Transactions on , vol.60, no.1, pp.400-414, Jan. 2012

[15] Nguyen, H., Prasad,N., Walker,C. and Walker,E. (2003). A First Course in FUZZY and NEURAL CONTROL. Chapman & Hall/CRC.

[16] NI Wireless Sensor Network (WSN) Starter Kit, http://sine.ni.com/nips/cds/view/p/lang/en/nid/206916, accessed May 2011.

[17] Paul, N.andDebono,C. (2008). An FPGA implementation of adaptive data reduction technique for WSN. InProc. of WICT 08 conference.

[18] Rappaport,T. S. (1996). Wireless communications: Principles and Practice(1st ed.). Prentice Hall.

[19] Recognizing Diseases Symptoms of Greenhouse Ornamental, http://www.greenhousemanagementonline.com/greehhouse-0211-recognizing-disease-symptoms-ornamentals.aspx, accessed June 2011.

[20] Ross,T. (2004). Fuzzy logic with engineering applications.John Wiley & Sons Ltd.

[21] Santini, S.andRömer,K. (2006). An Adaptive Strategy for Quality-Based Data Reduction in Wireless Sensor Networks. In Proc. of the 3rd International Conference on Networked Sensing Systems (INSS 2006). TRF, pp. 29-36, Chicago, IL, USA.

[22] Seah,W., Eu,Z. and Tan,H. (2009).Wireless Sensor Networks Powered by Ambient Energy Harvesting (WSN-HEAP) – Survey and Challenges. In Proc. ofVITAE 09.

[23] Shwe,H., hong,J., and Horiguchi,S. (2009). Energy saving in wireless sensor networks. Journal of Communication and Computer, ISSN 1548-7709, USA.

[24] Verdone,R., Dardari,D., Mazzini,G., and Conti A.(2009).Wireless sensor and actuator network technologies analysis and design. AP LCC, pp. 235-280.

[25] Wireless Sensor Network Simulator, http://www.djstein.com/projects/WirelessSensorNetworkSimulator.html, accessed April 2011.