Vol 7 Issue 2 Nov 2017

ISSN No : 2249-894X

Monthly Multidisciplinary Research Journal

Review Of Research Journal

Chief Editors

Ashok Yakkaldevi
A R Burla College, India

Ecaterina Patrascu
Spiru Haret University, Bucharest

Kamani Perera
Regional Centre For Strategic Studies, Sri Lanka

Welcome to Review Of Research

RNI MAHMUL/2011/38595 **ISSN No.2249-894X**

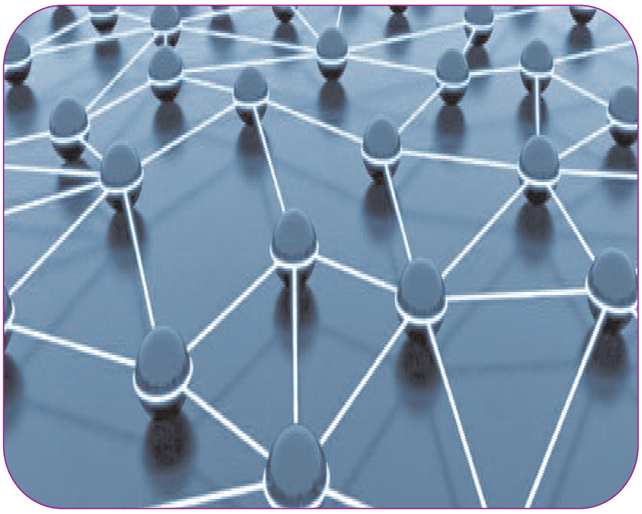
Review Of Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.oldror.lbp.world

REVIEW OF RESEARCH

ISSN: 2249-894X
IMPACT FACTOR : 5.2331(UIF)
VOLUME - 7 | ISSUE - 2 | NOVEMBER - 2017

A NOVEL SCHEME FOR ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORK APPLIED TO THERMAL COMFORT INDOORS



Thales Ruano Barros de Souza[1] **and Rogério Caetano**[2]

[1]Mestre em Engenharia Elétrica, pela UFAM.
[2] Doutor em Engenharia Elétrica, pela COPPE/UFRJ.

ABSTRACT

Wireless Sensor Networks can be used in several applications such as monitoring and study of thermal comfort in indoor environment. However, since the sensors nodes are battery powered, the cost of network maintenance can quickly exceed the cost of whole monitoring system. This work proposes and evaluates a routing protocol that take into account not only localization of sensor nodes but also the monitored variables, in this case, temperature and relative humidity to formation of clusters and selection of cluster heads. The k-means algorithm was used to group the correlated sensors nodes. Results show that the proposed scheme has a performance 50% higher than the state-of-the-art algorithms in relation to residual energy and network lifetime.

KEY-WORDS: *Wireless Sensor Netoworks; LEACH; Energy-efficiency;Thermal Comfort; Data Mining; k-means.*

I. INTRODUÇÃO

Wireless Sensor Networks (WSN) has received a lot of attention in the last few years. These networks consist in a number of sensor nodes with ability of sensing, processing and communication [1]. Due to its characteristics of auto-organization, low cost and rapid deployment, WSN are applied in many monitoring systems indoors to collect and process environmental data [2]. In this context, WSN have been used as a tool to study the thermal comfort. However, most studies only use WSN for continuous sensing of the environment, performing all processing on a base station, outside the network [2, 3].

The distributed nature of WSN can reveal microclimatic conditions that lead to thermal discomfort [3]. On the other hand, continuous sensing can be prohibitive in practical applications, due to excessive consumption of energy in the periodic transmission of sensed data. Moreover, the use of dense networks makes it necessary to use specific routing algorithms mainly due to problems that can take place in WSN, such as noise and packet collision, which leads to significant loss in both quantity and quality of the data. Routing is a critical task due to the resources constraints of WSN in terms of energy consumption, processing and bandwidth. For this purpose there are many routing protocols which employ the cluster structure, which allocate resources efficiently maximizing the network lifetime, avoiding problems as energy hole [4]. In this case, thestructure of sensor nodes is organized in clusters according to specifics requirements or metrics. Usually, each cluster includes a leader named as cluster head (CH) and other member nodes (MNs).The CHs perform the processing and the transmission of information outside de cluster in the direction of base station (BS), while MNs perform the sensing of the information and communicate only with its respective CH. The formation of clusters, especially the CHs selection, is the most critical task in management of the WSN. LEACH [5] is a pioneering approach of clustering routing in WSN. The main objective of LEACH is to use a stochastic model to select the CHs, thus the dissipated energy in the communication to BS is spread to all sensor nodes of the networks. The basic idea of LEACH has been inspiration for subsequent routing protocols. LEACH-C [6] proposes a centralized approach of clustering, while the protocol E-LEACH [7] improves the procedure of CH selection, by including residual energy as parameters of the mechanism of selection. TL-LEACH [8] introduced two level of clustering which leads to less average transmission distance, and less nodes are required to transmit far distances to the BS.W-LEACH [9] is able to handle non-uniform networks as well as uniform, in respect to energy consumption, increasing the average lifetime for sensor nodes. M-LEACH [10] protocol selects optimal path between the CH and the BS through other CHs and use these CHs as relay station to transmit data over through them. Thus CHs transmit data to the corresponding CH which is nearest to BS. Finally, this CH sends data to BS. T-LEACH [11] is a threshold-based CH replacement scheme for clustering protocols of WSN. It minimizes the number of CH selection by using threshold of residual energy.

In [2] is presented a centralized routing protocol that uses the k-means algorithm to monitor indoors. Nevertheless, a comparative analysis of the performance of the protocol was not performed. In [12] is proposed a dynamic clustering algorithm using genetic algorithm as method of CH selection. This algorithm uses the residual energy, which is the necessary energy to route a message to BS, and the number of CH to maximize the network lifetime. The purpose of these last works is to minimize the distance between MNs and their respective CH, to reduce the intra cluster cost communication.

The problem to be dealt with in this paper is how to extend the lifetime of a WSN deployed indoors to monitor environmental variables in case of user's thermal comfort applications. In this research we are interested in extend the lifetime of the network to avoid constants maintenance which can exceed the cost of the whole system in a short period of time. We minimize the energy consumption not only reducing the distance between MNs and their respective CHs but also by reducing the quantity of intra cluster transmissions through the k-means algorithm. The sensor node elected as CH is the sensor that better represents the cluster to which it belongs. Experimental results show that the proposed scheme presents better performance when compared to LEACH [5] and [13], using residual energy and network lifetime metrics.

The rest of this paper is organized as follow. Section 2 presents the related researches, and Section 3 introduces the proposed method. The simulation results and evaluation of the proposed scheme are presented in Section 4. Finally, Section 5 concludes the paper and suggests the future work.

II. BACKGROUND

In this section the existing representative cluster-based routing protocols proposed for WSN including LEACH [5] and a K-means Approach [13] are discussed.

A. LEACH

The basic idea of LEACH is to select CHs by rounds, thus the energy consumed in communication with BS is spread throughout the network. The operation of LEACH occurs in several rounds, where each round is separated in two steps, set-up phase and stationary steady. During set-up phase the clusters are organized. While in the steady phase the data are delivered to BS as we can see in Figure 1. Set-up phase can be divided in three steps: advertisement, cluster set-up and schedule creation.

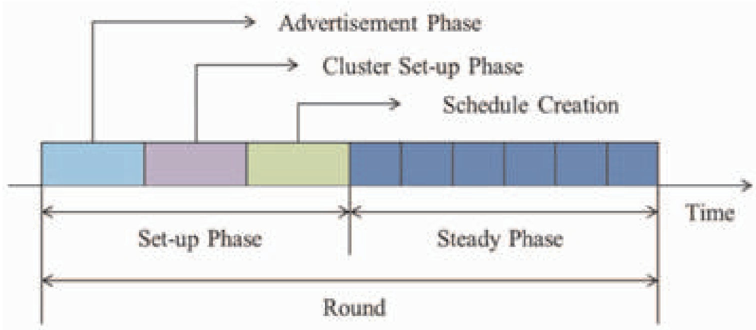


Figure 1.The structure of one round of LEACH protocol [9].

During the set-up phase, each sensor node decides if it will become a CH or not in actual round. This decision is based on the percentage of clusters suggested to the network and the number of times that each sensor node was elected as CH at that given moment. That decision is taken from the time at which each sensor node randomly chooses a number between 0 and 1. Sensor nodes become CH in the actual round if the number is smaller than the threshold:

$$T(n) = \begin{cases} \frac{P}{1 - P\left(r \bmod \frac{1}{P}\right)}, & if\ n \in G \\ 0, & otherwise \end{cases} \qquad (1)$$

Where P is the desired percentage of CHs, r is actual round and G is the set of nodes that were not elected as CHs in the last 1/P rounds. When one sensor node is elected as CH, it broadcasts a message to the other nodes. According to the received signal strength, other node decides which group to enter in the actual round and sends message of association to the closer CH. Rotation of CHs occurs in each round through the generation of a new set-up phasein order to evenly distribute power consumption among sensor nodes. During steady phase, sensors nodes collect and transmit data to the CHs. CHs perform compression of the data and deliver an aggregated packet to BS. After a period of time, that is determined a priori, the network returns to set-up phase and it performs a new round of selection of CH. Figure 2 shows a basic topology of LEACH.

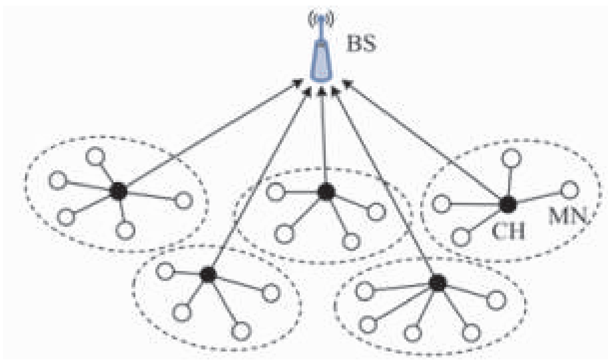


Figure 2.Basic topology of LEACH.

However, LEACH has some disadvantages, like a single hop inter cluster routing method, directly from CHs to BS, what is not applicable for large scale network. Moreover, as the protocol does not take into account residual energy, LEACH can't ensure load balancing in the case of sensor nodes with different quantity of initial energy. Finally, as election of CHs is performed in term of probability, it is difficult to obtain a uniform distribution of CHs in the entire network.

B. K-means Approach

Unlike probabilistic approach of LEACH, in [13] a deterministic approach is developed, using k-means algorithm, to maximize network lifetime of WSN. The k-means algorithm is used to form clusters so that the distance between sensors nodes and CH become minimal. That method allows minimize the energy consumption of sensor nodes sending data to CH. As consequence, the network lifetime is prolonged.

Such a method uses the k-means algorithm to form clusters based on Euclidian distance between sensor nodes. After clusters formation, an ID is attributed to each sensor node of a cluster according to distance of centroid, so that closer sensor nodes receives lowest ID. Figure 3 show how the IDs are attributed. ID indicates the order to elect CHs. Therefore ID plays an important hole on the selection of CH. Besides that, residual energy is checked in each round for network connectivity retention. If the CH's energy is less than a threshold predetermined, the next sensor node is select as new CH. The new CH elected notify others nodes about the change of CH.

In each round, k-means approach adopts single hop routing protocol in the transmission of data of CHs to BS, as shown in Figure 2. Collected data are processed in BS.

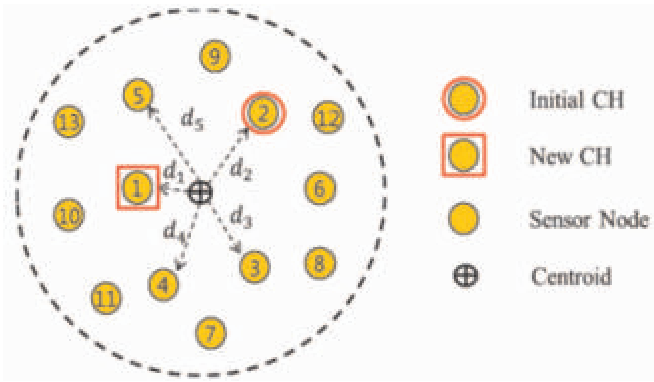


Figure 3.Ordering of the nodes with ID number [9].

However there is a strong dependency of ID's sensor nodes, which is used as selection criterion of CHs causing more overhead of control messages.

III. PROPOSED METHOD

In this section we present the proposed CH selection method which employs the k-means algorithm. This method take into account not only localization of sensor nodes but also the monitored variables, in this case, temperature and relative humidity to formation of clusters and selection of CH.

A. Network Model

In this section, we will define network topology and model of energy consumption.

The network topology adopted in this work was the same of the experiment realized into Berkeley's laboratory [14] as shown in Figure 4. There the network is composed by 54 sensor nodes distributed in the laboratory. Each node collects temperature and relative humidity of the environment.

In this paper, the model of propagation adopted was multipath model, because it is an indoor application. In indoors there is a lot of transmitted signal's degradation due to obstacles and noise of the environment. Thus, the model of energy consumption adopted in this paper is based on transmissions and receptions like as can be seen in Eq. (2) and (3):

$$E_{Tx}(k,d) = E_{elec} * k + \varepsilon_{mp} * d^4 \quad (2)$$

$$E_{Rx}(k) = E_{elec} * k \quad (3)$$

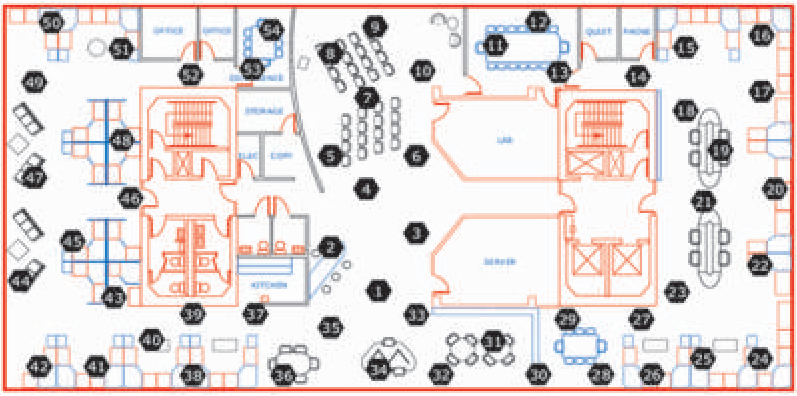


Figure 4.Wireless sensors installed in the Intel Berkeley Research Lab [14].

Where E_elec is the necessary energy to process 1 bit of data with the electronic circuit and ε_mp is the consumed energy by the transmission of 1 bit of data to achieve an acceptable error rate in the multipath model. Table 1 shows values of parameters to the simulation.

Table 1. Simulation parameter

Parameters	Values
Network Size	(40.5m x 31m)
Location of BS	(24.5m:12m)
Number of nodes	53
Initial Energy	0.5J
E_{elec}	50nJ/bit/m^2
ε_{mp}	0.0013pJ/bit/m^2
Sizeof data packet	6400 bytes
Energy taken for aggregation (E_{DA})	5nJ/bit

To define the optimal amount of groups we use silhouette function [15]. This function measure the similarity among collected data and localization of sensor nodes inside and outside the clusters. This measure vary between +1, indicating points that are tightly grouped, through 0, indicating points that are in the periphery of the clusters, until -1, indicating points that probably were assigned to wrong clusters. Figure 5 shows the silhouette values for 5 clusters. As can be seen there is much more silhouette values greater than 0.5 than silhouette values lesser than 0.

Therefore to choose the quantity of clusters, we vary k among 3 and 10 and we defined a threshold described in Eq. (4):

$$y = sum\_positive - sum\_negative \quad (4)$$

Where sum_positive represents the sum of values whose value attributed by silhouette function is greater than 0.5 and sum_negative represents the sum of values whose value attributed by silhouette function is lower than 0.

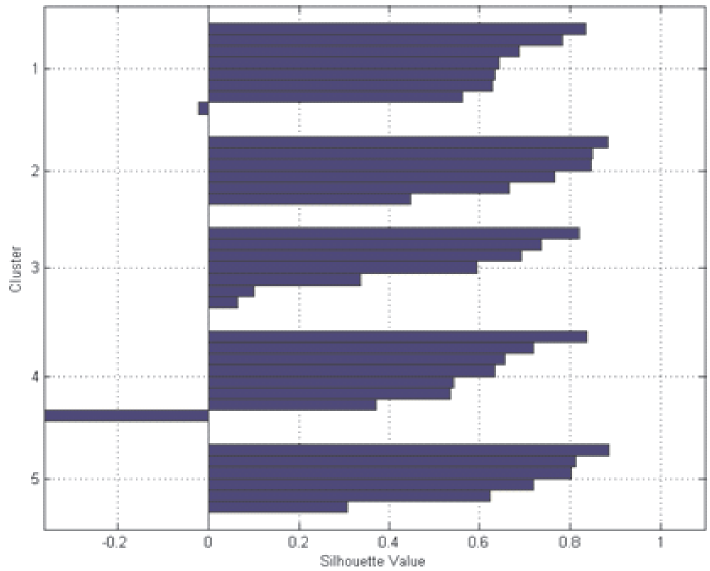


Figure 5 Silhouette values for 5 clusters.

Figure 6 presents the result of Eq. (4) for k varying among 3 and 10. We can realize that for k=5 the result presents maximumvalue, indicating that this is the better choice for this scenario.

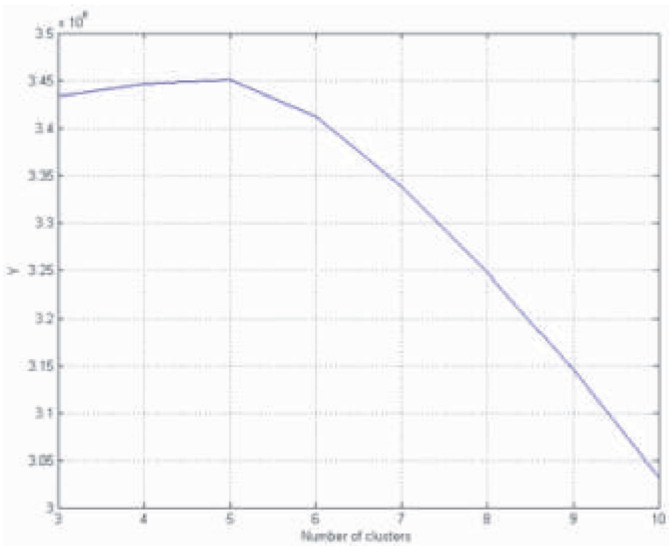


Figure 6.Defining the number of clusters.

B. Routing Protocol

Like LEACH, the proposed scheme is composed by set-up phase and steady state phase as shown in Figure 1. In the set-up phase there is the exchange of control messages to select CHs, the announcement of CH to network and the input of each sensor node in a cluster. On the other hand, in steady phase each sensor node collects periodically information about the environment and transmits this information to CH, which in turn, transmits the information to BS.

Step 1: Initial clustering

Initially we use the localization of each sensor node as input of the k-means algorithm. Therefore the sensor node closest of group's centroid is selected as CH. BS broadcast information about sensor node chosen as CH so that they are aware of your function in the network. Thereafter, each CH performs broadcasting so that each sensor node decides to which CH to associate. Posteriorly each sensor node collects environment information, in this case temperature and humidity, and sends to respective CH, which in turn route the information to BS.

Step 2: Re-clustering

In this stage occurs clustering of k-means algorithm, but this time the inputs are: localization (X, Y), temperature and relative humidity of each sensor node. Thereat, CH's selection occurs through the minimum distance in relation to the four dimensions. Thus, in steady phase there is no communication among each sensor node and CH, saving the energy consumed in intra cluster communication. Leader informs to another sensor nodes about change of CH. For each round, the proposed scheme adopts a single hop routing protocol, both in communication intra cluster and inter cluster, as shown in Figure 2.

Step 3: CH selection

Algorithm CH Selection describes the process of CH's selection, where function MinDist() seek sensor node with closest reading to centroid of each cluster, as shown in Figure 7. Thus, the CH selected is the sensor node that best represents the cluster in relation to temperature and humidity measurements. Function InformMsg() is responsible to notify another sensor nodes about change of CH. After that, only CHs send data do BS. At this point the energy consumed in the intra cluster communication is saved.

Algorithm CH Selection

Procedure	
1.	Input: number of clusters, temperature and humidity readings of each sensor node
2.	Clusters formation through k-means algorithm
3.	CH = MinDist()
4.	All nodes<- InformMsg()
5.	Send data to BS

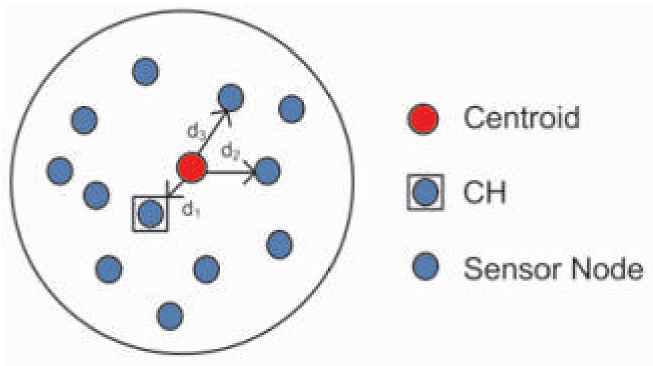


Figure 7.CH selection.

The proposed method flowchart can be seen in Figure 8. We can realize that the algorithm is composed by two phases, set-up phase where there is clusters formation and selection of CHs, and steady phase where CHs send information to the BS. As all the process is based in rounds, after a period of time determined a priori, set-up phase will happen again. But if the cluster were formed, there is no more initial clustering. Instead, there is re-clustering, where the k-means is performed with localization of the sensor node and the new values of temperature and relative humidity measured by the network.

IV. PERFORMANCE EVALUATION

In this section, we present the evaluation of the proposed method through Matlab simulation and we compare with two routing protocols. The routing protocols selected to comparison were LEACH and the proposed in [13]. The metrics to be evaluated are energy efficiency and lifetime of the network.

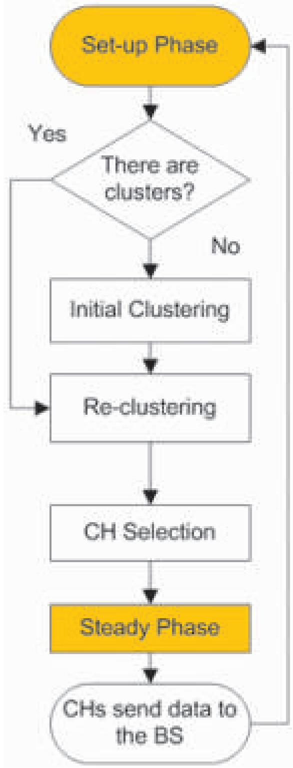


Figure 8.Flowchart of the proposed scheme

The performance evaluation was carried out by simulation. In this case, we used the database available at the Berkeley's laboratory [14]. This database has more than 2.3 million of readings of 54 sensor nodes. Figure 9 shows data volume that each sensor node generated in the period of 28 of February and 5 of April of2004. As can be seen, nodes among 21 and 31 collected more data than the rest of the network.

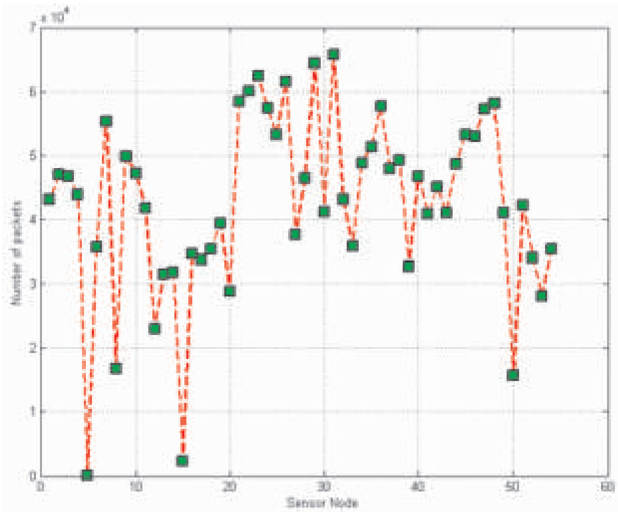


Figure 9.Data volume generated by sensor nodes.

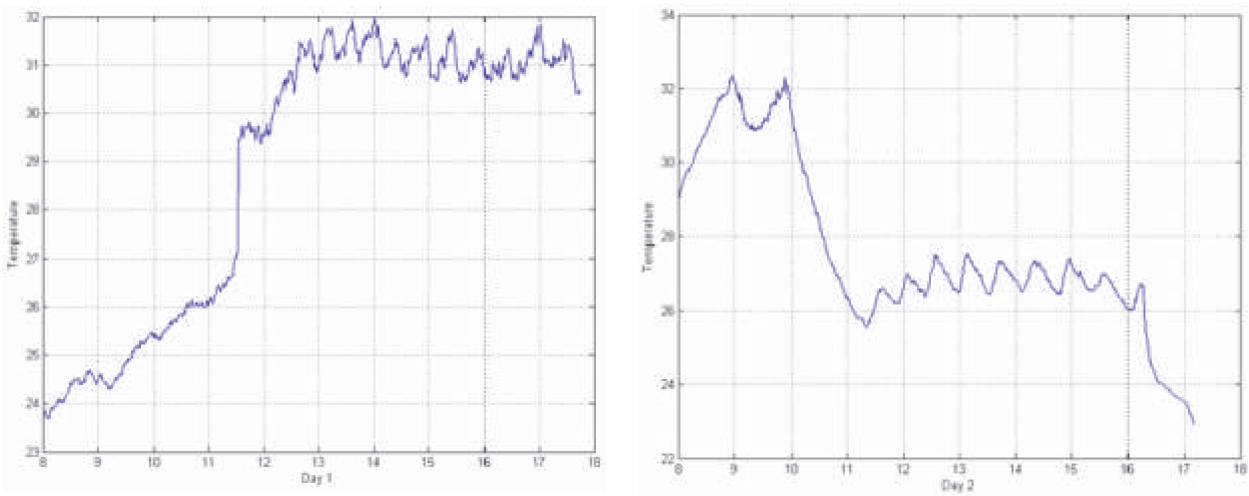


Figure 10 Temperature distribution on 1st and 2nd March.

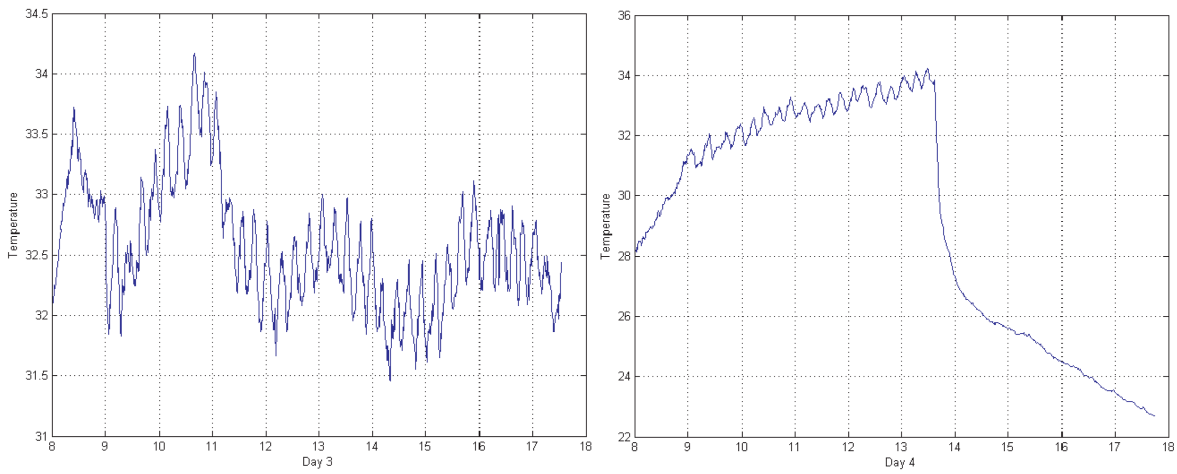


Figure 11 Temperature distribution on 3rd and 4th March.

It is very hard to estimate the temperature distribution of a place in a day. For example, given four days, March 1-4, we can find the temperature distributions collected by sensor node 22 are different (Figure 10 and 11). They showed different overall patterns. For example, the temperature on 1st March is increasing dramatically from 8:00 am to 14:00, whereas the temperature on 2nd March changes with a different pattern comparing to that on 1st March.

As the data correlation structure of the data will be lead into consideration to form clusters and selection of CH, it was necessary pretreatment data to remove invalid data and outliers, which are values that are far away when compared with most data. Methodology of treatment of data was similar of the performed in[3].

The following steps are conducted to remove observations that are not valid for various reasons.

• Invalid ID: ID of the sensor nodes are labeled from 1 to 54 in the dataset, that is $1 \leq ID \leq 54$.
• Invalid temperature: Assume the temperature in the buildings ranges from 10 to 40°C.
• Invalid humidity: As indicated in the website [14], the humidity falls in 0-100%. All of the observations satisfy this condition, and hence no observation is removed from the dataset.
• Invalid voltage: Set $2 \leq voltage \leq 3$.

Having conducted the above cleansing procedures, we found the data collected from 26th March and 5th April 2004 has been removed. We have an impression that the original dataset has large amount of invalid observations, or in data mining terminology, the data is very dirty.

All of the above cleansing procedures check the validation of one variable. As the number or removed observations is so big. As the number of removed observations is so big, we doubt there might be outlying observations, or called outliers, in the dataset.

Figure 12 shows the number of the observations collected within the rest days, from 28th February to 25th Match (denoted as days 1-27 in the figure). From Figure 9, the last 2 days – days 26 and 27 (i.e., 24th and 25th March) - collects a small amount of observations.

• Observations collected on 25th March. There only 218 observations collected on the last date (25thMarch), and all of these observations are from sensor node 9. As all of these observations are from one sensor node, which is not representative enough, we remove them from the dataset. Having been removed these observations, the dataset has 774,446 observations left.

• Observations collected on 24th March. There 1827 observations on 24th March. As they are from a number of different sensor nodes, we keep them in the dataset.

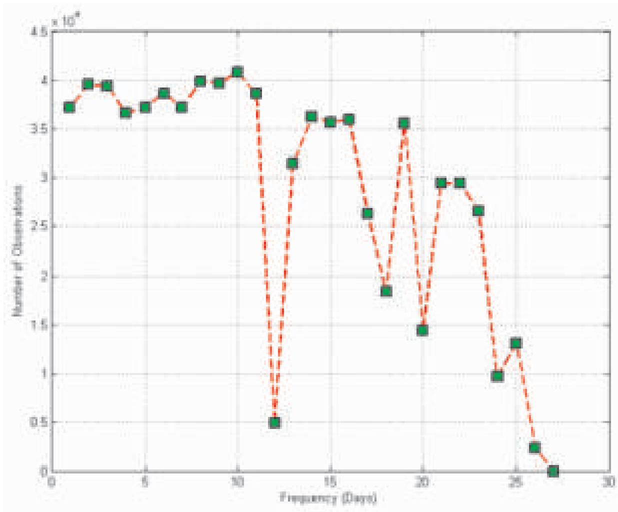


Figure 12 Number of observations per day.

There might be outliers within the rest of observations. We can use clustering algorithm to detect the possible outliers. Among the clustering algorithms, the k-means algorithm is usually used for large datasets. As our dataset is so big, we select the k-means clustering algorithm to detect outliers. The k-means clustering algorithm has been widely used in data mining and statistical data analysis [16].

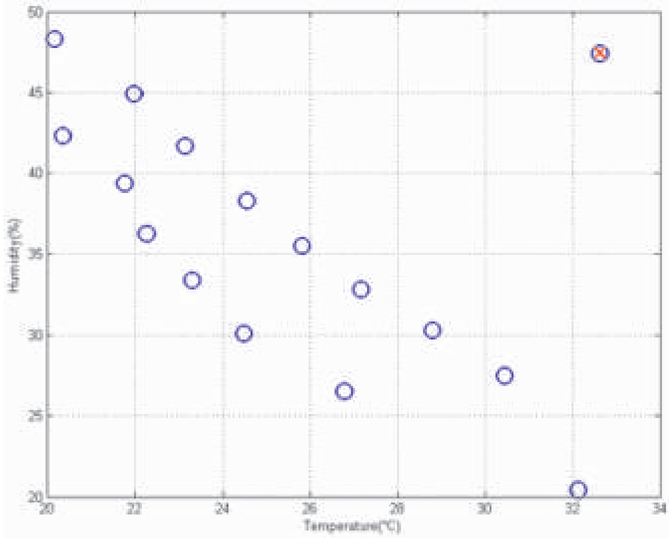


Figure 13.Centroids of 16 groups.

K-means algorithm was used to identify outliers. When K=16, both temperature and humidity are abnormally big in the smallest cluster. In Figure 13 can be seen the centroids of 16 groups, including centroid that corresponds to the group that was classified as outlier marked in red. Samples that belong to this group were removed, remaining 738,604 observations. Most of the discarded samples were collected in the last days, when energy level of the sensor nodes was already low.

Lifetime of the network is defined as the number of rounds wherein energy of all sensor nodes ends. Figure 14 shows residual energy of the network as the round proceeds. As it can be seen, when compared to proposed scheme, LEACH is 50% worse and routing protocol of [13] is 30% worse. We can observe that after 1000 rounds there is low energy remaining on other protocols. However network still have residual energy at the round 2000 in the proposed scheme.

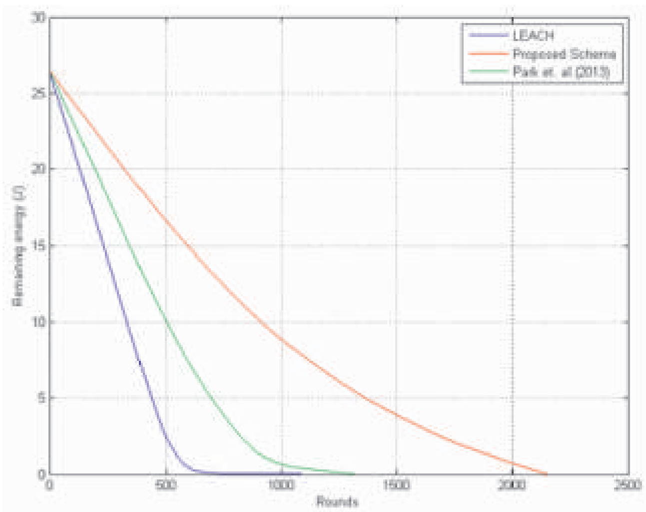


Figure 14.Residual energy of the network per round.

The number of live nodes is other form to analyze lifetime of the network. As it can be seen in Figure 15, after round 1323 there is no sensor node alive on the other protocols. On the other hand, in the proposed scheme after round 2000 there is still alive node sensors. It is clear that the lifetime of the proposed scheme is much longer than that of LEACH and the algorithm proposed in [13]. This is because the intra communication overhead of our proposal is lower than other protocols. Note that the proposed scheme of this paper forms clusters such that the distances between the CH and the member nodes are minimized as much as possible and we select as CH the sensor node that best represent the cluster in relation to temperature and humidity.

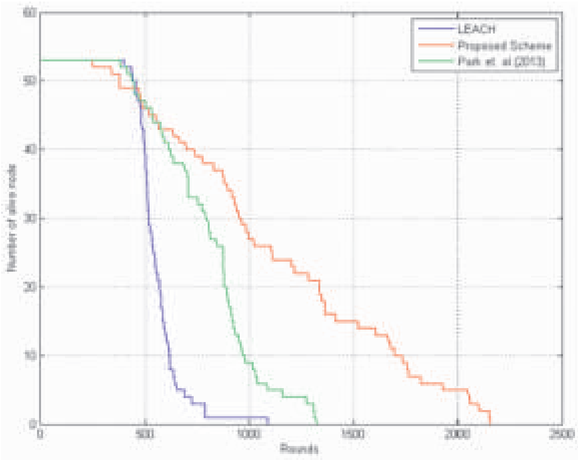


Figure 15.Number of live nodes per round.

V. CONCLUSION AND FUTURE WORKS

A major concern in applications of RSSFs is the network lifetime, even in indoors applications, where the maintenance cost can overcome quickly the cost of implantation of the entire monitoring system. To solve this problem, routing protocols have been proposed. Nowadays, most of the proposed schemes seek to reduce distance among sensor nodes to respective CHs, as proposed in [13], through unsupervised learning algorithm, as k-means. Therefore energy consumption of the intra cluster communication is reduced.

This paper proposes not only reduce distance of communication intra cluster, but also choose sensor node that best represent each group as CH. Thus there is a reduction of intra cluster communication, consequently reducing energy consumption. Performance evaluation results show that the proposed scheme reduces energy consumption up to 50% when compared to others routing protocols of the literature, as LEACH. As future work, we propose the development of an event oriented model for detecting abrupt variations (Figure 10 and 11) in temperature and humidity, which can go through the development of filters on the sensor nodes.

ACKNOWLEDGMENT

The authors acknowledge the support granted by the Center of Excellence in Science and Technology Program (PRONEX/FAPEAM/CNPq) through process 01135/2011.

REFERENCES

[1] I.F. Akyilidiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," Computer Networks, vol. 38, no. 4, pp. 393-422, 2001.

[2] X. H. Li, K. L. Fang, L. Z. He, "A clustering algorithm based on k-means for wireless indoor monitoring system," International Conference Information Technology and Computer Science, 2009.

[3] S. Wu, D. Clements, "Understanding the indoor environment through mining sensory data – A case study," Energy and Buildings, 2007.

[4] J. Li, P. Mohapatra, "Analytical modeling and mitigation techniques for the energy hole problem in sensor networks," Pervaisve Mobile Comput. 2007, 3, 233-254.

[5] W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishan, "Energy efficient communication protocol for wireless microsensor networks," In Proceedings of the 33rd Annual HawaiiInternational Conference on System Sciences, Maui, HI, USA, 4–7 January 2000; pp. 10–19.

[6] W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Trans. Wirel. Commun. 2002, 1, 660-670.
[7] X. Fan, Y. Song, "Improvement on LEACH protocol of wireless sensor network," in Proceeding of International Conference on Sensor Technologies and Applications, Valencia, Spain, 14-20 October 2007, pp. 260-264.
[8] V. Loscri, G. Morabito, S. Marano, "A two-level hierarchy for low-energy adaptive clustering hierarchy," in Proceedings of the 2nd IEEE Semiannual Vehicular Technology Conference, Dallas, TX, USA, 25-28 September 2005; pp. 1809-1813.
[9] H. M. Abdulsalam, L. K. Kamel, "W-LEACH: Weighted Low Energy Adaptive Clustering Hierarchy aggregation algorithm for data streams in wireless sensor networks," in Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW), Sidney, Australia, 14 December 2010, pp. 1-8.
[10] M. Xiaoyan, "Study and Design on Cluster Routing Protocols of Wireless Sensor Networks," Ph.D. Dissertation. Zhejiang University, Hangzhou, China, 2006.
[11] J. Hong, J. Kook, S. Lee, D. Kwon, S. Yi, "T-LEACH: The method of threshold-based cluster head replacement for wireless sensor networks," Inf.Syst. Front. 2009, 11, 513-521.
[12] M. A. Mehr, "Design and implementation a new energy efficient clustering algorithm using genetic algorithm for energy efficient wireless sensor network," World Academy of Science, Engineering and Technology, 52, 2011.
[13] G. Park, H. Kim, H. W. Jeong, H. L. Youn, "A novel cluster head selection method based on k-means algorithm for energy efficient wireless sensor network," 27th International Conference on Advanced Information Networking and Applications Workshops, 2013.
[14] Intel Berkeley Research Lab, http://db.lcs.mit.edu/labdata/, 2004.Acesso em: 10 jun. 2013.
[15] L. KaufMan, P.J. Rousseew, "Finding Groups in Data: An Introduction to Cluster Analysis," Hoboken, NJ: John Wiley & Sons, Inc., 1990.
[16] I.H. Witten, E. Frank, "Data ming – Practical Machine Learning Tools and Techniques," Morgan Kaufmann Publishers, San Fracisco, CA, 2005.