Efficient Hierarchical Clustering Routing in Wireless Sensor Networks

Rais Amine[1,*], Bouragba Khalid[1] and Ouzzif Mohammed[1]

[1]*Laboratory RITM, ENSEM, Hassan II University of Casablanca, Casablanca, Morocco*
amine.343@hotmail.fr

Abstract

The routing protocol allows routing data from source to destination through intermediate nodes up to the sink. In the hierarchical routing based on the clustering, every cluster head node forms its cluster with their neighbor nodes.The cluster heads are responsible of their clusters. In fact they coordinate between nodes in the cluster for intra-cluster communication and route data to the other clusters for inter-cluster communication. In this article we propose a centralized and periodic protocol named Efficient Clustering Routing Protocol (ECRP). Our protocol is based on a modified version of CLARANS algorithm; it takes into account the location and the remaining energy of the sensor nodes to divide the sensed area to disjoint and balanced clusters.The simulation results show that the proposed protocol is more efficient than the existing hierarchical protocols in terms of packet delivery ratio, and energy consumption.

Keywords: Clustering; routing; wireless sensors networks; CLARANS; energy consumption; latency

1. Introduction

The routing in Wireless Sensor Networks(WSN) consists of conveying the gathered data from the sensors to the sink but this operation is very energy demanding. The communication between the nodes is the process that consumes the most of the energy; about 70% of the total energy consumed by the nodes of the network. However, a good a routing policy should be adopted to save the limited resources of the nodes and therefore expand the lifetime of the network. This goal could be achieved by using a robust and scalable routing protocol that takes into consideration the residual energy, and the optimal routing itinerary of data transmission between the sensor nodes.

The WSN can be classified according to different criteria; network topology remains the most common criterion which subdivides WSNs into 3 sub categories: plate routing protocol, hierarchical routing protocol, and the geographical routing protocol.

In this article we will focus on the hierarchical routing protocol based on the clustering and we will present a new one dubbed Efficient Clustering Routing Protocol (ECRP) that makes use of the CLARANS algorithm[1] in order to partition the network intobalanced clusters and to determine a cluster head (medoid) for each cluster.

The rest of this paper is organized as follows. Section 2 presents hierarchical clustering routing protocol in WSN. In Section 3 we present the radio model of communication. Section 4 describes ECRP protocol. The simulation and results are proposed in Section 5 and we conclude this paper in Section 6.

2. Hierarchical Clustering Routing in WSN

The clustering is the process that allows the partitioning of nodes into groups names clusters. Each cluster contains a certain number of member nodes and one cluster head

node. The clustering technique is very promising because it offers a lot of benefits to WSN. In fact, the organizing of the sensors into clusters allows creating a logical hierarchy of the nodes. It also helps reducing the complexity of the routing and optimizing the data aggregation. This technique also resolves the scalability problem and finally it allows minimizing the energy consumption, thus, expanding the lifespan of the network. Amongst the most common hierarchical protocols we can cite: leach, leach-c, pegasis, and H-pegasis.

Heinzelmen *et al.*, propose in [2] the leach protocol to divide the sensors of the network into two levels, they divide the sensors of the network into two levels: member nodes and cluster head nodes, leach uses the powerof the receivedradio signal to create clusters, every sensor node decide autonomously to be cluster head node or a member node, it takes a randomized value V between 0 and 1, if $V<T(n)$the sensor node will be designed as a cluster head, with $T(n)$ is a threshold determined by:

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

Where P is the desired percent of cluster head in the network, r is the number of round and G is the set of sensor nodes that have not become cluster head during the last $\frac{1}{P}$ rounds.

The Leach protocol is broken into rounds, every round is composed from 2 phases: set up phase and steady state phase. In the set up phase the clusters will be created, the sensor nodes selected as cluster heads broadcast an advertisement message to the rest of sensor nodes, which decide the cluster head to join for the current round, the nearest cluster head is selected based on the received signal strength of the advertisement.

In the steady state phase the member nodes may communicate with their cluster head nodes using the TDMA schedule, a time slot is allocated for every sensor node to transmit data to the cluster head.

The arbitrary selection of CHs in leach causes an unbalanced distribution of CH, dissipates the energy of the sensors and consequently reduces the lifetime of the network, for that a centralized version of leach named Leach-c is proposed in [3], a global view of the network is formed and the mission of cluster creation is taken by the sink.

PEGASIS is proposed by Lindsey et al. in [4], which consists of creating a long chain of nodes which allows each node to communicate only with the closest node. The communication with the sink is organized in rounds allowing only one node to directly communicate with the sink. PEGASIS is known for its low dissipation of energy but suffers from long date delivery time. It also lacks reliability inevery iteration when the cluster head is down. An enhanced version of PEGASIS: hierarchical PEGASIS was developed in [5] to address these issues by adopting alternative ways of communication with the sink in the case of geographically remote nodes. There are also many hierarchical protocols that give good performances like: HEED [6], EEHC [7] and EE-LEACH [8].

3. Radio Model of Communication

The transmission of data may consume a big amount of energy and exhaust the battery power of the sensor nodes. The client/server architecture is not efficient to save energy in WSN especially for networks with high traffic,the sensor nodes close to the sink are critical nodes; they gather their own data and transmit the data of the other nodes to the sink. The depletion of energy in these sensor nodes can collapse the whole network and routing data towards the sink becomes impossible. In order to resolve this problem a good policy of energy management must be followed.

According to Heinzelman *et.al.*, [2] to send a message of *m*bits to a sensor node from *d* distance the radio energy used is:

$$E_{TX}(m,d) = E_{elec} * m + E_{amp} * m \qquad (2)$$

To receive this message the energy dissipated would be calculated as:

$$E_{RX}(m) = E_{elec} * m \qquad (3)$$

Eelec is the energy consumed to run the transmitter or receiver circuitry and *Eamp* is the energy consumed to run by the transmission amplifier, this former will be calculated by:

$$\begin{cases} Eamp = \theta_{fs} * d^2 & \text{if } d \le l_0 \\ Eamp = \theta_{mp} * d^4 & \text{if } d > l_0 \end{cases} \qquad (4)$$

Where $l_0 = \sqrt{\dfrac{\theta_{fs}}{\theta_{mp}}}$, θ_{fs} and θ_{mp} are the radio amplified consumed energy for free space propagation and multipath fading channel model, according to Heinzelmen *et al.*, in [3] the value of θ_{fs} and θ_{mp} are respectively 10pJ/bit/m^2 and 0.0013pJ/bit/m^4.
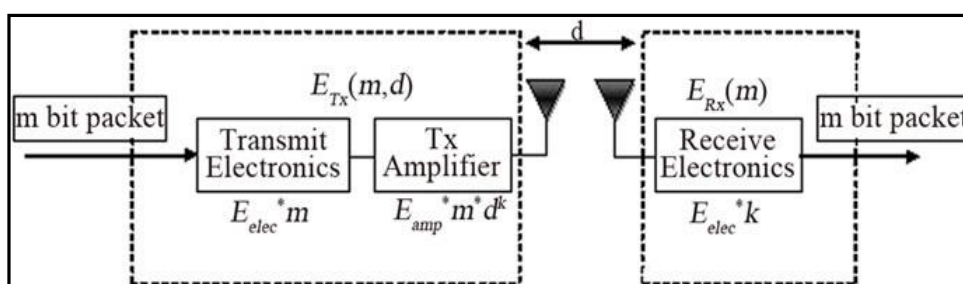

Figure 1. Radio Model Energy Consumption

4. Description of ECRP Protocol

The over consumption of energy due to the re-election procedure of the CHs as well as the re-arrangement of the clusters makes the existing clustering algorithms non suitable to WSN for which the energy consumption is a major constraint. None of these algorithms can ensure a fair distribution of cluster heads in space. In fact, these algorithms don't take into account the geographical distribution of nodes in order to choose the cluster heads. In the existing clustering algorithms, the nodes join the nearest cluster head without any control over the size of the cluster, which can result into overcrowded or under crowded clusters.

Our protocol include two phases: set up phase followed by a steady state phase. The set up phase is running in the level of the sink which has an overview of the network to create clusters and determine the cluster head nodes. The steady state phase is where the data transmission begins between sensor nodes and their cluster head nodes.

4.1. Set up Phase

Initially, all the nodes of the network are in listening mode. The sink sends out requests to the nodes in order to obtain their coordinates and the remaining energy of each sensor in the sensing area. The sensor nodes are aware of their location using trilateration and triangulation techniques [9, 10, 11] and no GPS receiver is required for every sensor since the GPS system is costly. The signal reception is energy consuming and may not function where physical obstacle exist.

We use the clarans algorithm to create clusters and to choose the cluster-heads which is the medoid of each cluster, it is approximately the sensor node located in the center of every cluster. Clarans try to minimize the within-cluster scatter and find the best group assignments for each node.

The duration of the execution of our algorithm to get the final clustering depends on the initial clustering medoid nodes; we have to choose the suitable medoid nodes to begin the set up phase.

Let M the medoid of all the nodes in the network. To find this node we have to calculate the average distance $\overline{X}_i$ between the medoid and the other nodes of the network and deduce the minimal value of $\overline{X}_i$ ($\text{Min}\overline{X}_i$).

$$\overline{X}_i = \frac{\sum_{j=1}^{n} dij}{n} \tag{5}$$

With n is the number of sensor nodes in the network and dij is the distance between the nodes i and j. It will be calculated as:

$$dij = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{6}$$

(x_i, y_i) and (x_j, y_j) are the coordinates of the nodes i and j, thus we deduce M with $\overline{X}_M$ equal to $\text{Min}\overline{X}_i$, finally the coordinates of the initial medoid nodes will be calculated as:

$$\begin{cases} mdi_x = \overline{X}_M \times cos(\frac{2\Pi}{k} \times (i-1)) + M_x \\ mdi_y = \overline{Y}_M \times sin(\frac{2\Pi}{k} \times (i-1)) + M_y \end{cases} \tag{7}$$

With k is the number of clusters in the network and i=1, 2, 3,…, k.

The medoid node i is the closest sensor node to the coordinate (mdi_x, mdi_y), thus we can specify the initial k-medoids nodes.

Heinzelmen et al define in [3] the value of k with the following formula:

$$k = \frac{\sqrt{n}}{2\Pi} \times \sqrt{\frac{\Theta_{fs}}{\Theta_{mp}}} \times \frac{L}{d_s^2} \tag{8}$$

L is the side of the sensed square field and d_s is the average distance between sensor nodes of the network and the sink.

The objective of our protocol is to choose the most centrallized node in every cluster as a medoid to be the cluster-head and to minimize the average distance between medoids and nodes of its clusters:

$$minavgs \sum_{i=1}^{k} \sum_{j \epsilon Ci} dij \tag{9}$$

With i is a medoid, j a node in the cluster Ci and dij is the distance between i and j.

After the choice of the initial medoid nodes, every non selected node j will be assigned to the nearest medoid m:

Algorithm1. The grouping algorithm for the set of non selected nodes (nSN)

> **For** each node j in nSN **do**
>> initialise min to a large value
>> **For** i from 1 to k **do**
>>> Calculate d_{ij} between the node j and the medoid i
>>> **If** $d_{i,j}$<min **then**
>>>> min← d_{ij};
>>>> $m \leftarrow$ i;
>>> **End if**
>> **End For**
>> $C_m \leftarrow C_m$+ j;/*the node j will be assign to the cluster represented by the medoid m*/
> **End For**

A swap between a selected medoid node i and a non representative node h will be necessary to find the best set of medoids for our network and to improve the quality of clustering,$C_{i'}$ is the cluster represented by the node i', with i' is the second most similar medoid to the node j, to calculate the cost S_{ih} of this operation we will have 4 cases for every non selected node j.
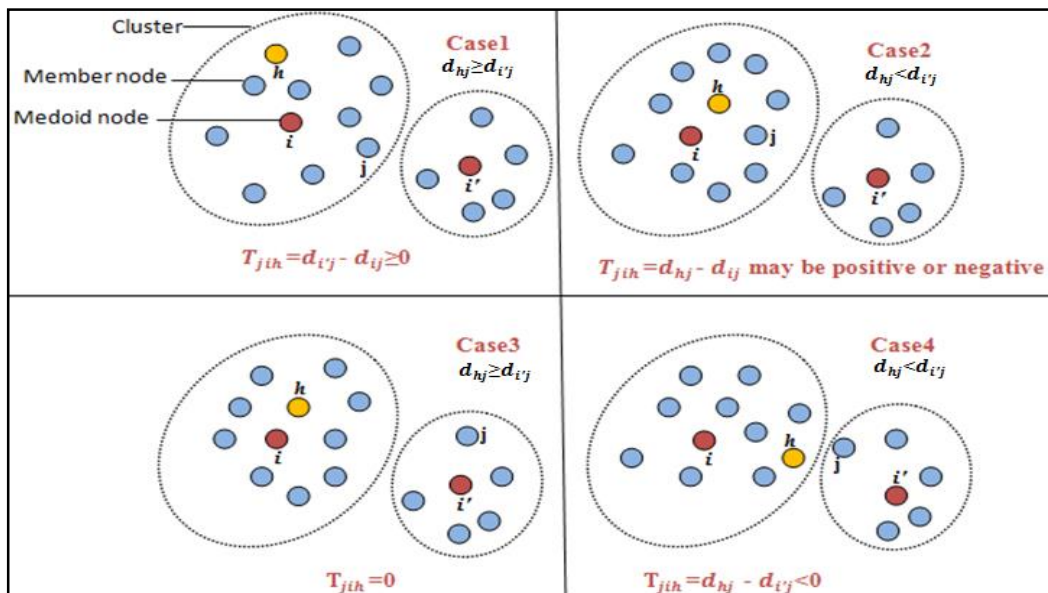

Figure 2. Cases of Swapping Medoid Node with a Non Representative Node

Algorithm2. Calculation of the swapping cost

Initialization

$S_{ih} \leftarrow 0$;

$C_h \leftarrow \emptyset$;/*C_h is the cluster represented by the node h*/

Begin

For each node j in nSN **do**

  If j in C_i **then** /*if the node j belong to the cluster C_i */

    If $d_{hj} \geq d_{i'j}$ **then** /*case1*/

      $C_{i'} \leftarrow C_{i'}$+j;/*the node j will be assign to the cluster $C_{i'}$*/

      $T_{jih} = d_{i'j} - d_{ij}$; /*with T_{jih} is always positive*/

    Else /*case2*/

      $C_h \leftarrow C_h$+j;/*the node j will be assign to the cluster C_h*/

      $T_{jih} = d_{hj} - d_{ij}$;

    End if

  End if

  If j in $C_{i'}$ **then**

    If $d_{hj} \geq d_{i'j}$ **then**/*case3*/

      $T_{jih} = 0$;/*the node j stay in the same cluster $C_{i'}$*/

    Else /*case4*/

      $C_{i'} \leftarrow C_{i'}$-j;/*the node j will quit the cluster $C_{i'}$*/

      $C_h \leftarrow C_h$+j;/*the node j will be assign to the clusterC_h*/

$T_{jih}=d_{hj}-d_{i'j}$; /*with T_{jih} is always negative*/

  End if

  End if

 $S_{ih} = S_{ih} + T_{jih}$

 End for

 Return S_{ih};

End

We make some modification in the clarans algorithm to be suitable with the WSN environment,we divide our network into uniform and distributed clusters, and we search the cluster head node for every cluster. It will be in the center of each cluster and take the name of medoid. This algorithm allows the node to arange itself into local clusters taking into consideration the distance between the nodes and the remaining level of energy in each node. The clustering of the neighboring nodes into the same cluster allows a better agregation and a low level of data exchange within the same cluster.

Algorithm3. Modified version of clarans algorithm

Initialization

Based on the equation (7) we calculate the set of initial medoid nodes I;

j← 1; // counter of neighbors

Begin

  Assign I as the current set of medoids C;

  While j >Maxneighbor **do**

    Select arbitrary a neighbor N of C;

    If Cost(C,N)< 0 **then**

      Assign N as the current set of medoids C;

      j = 1;

    Else j++;

    End if

  End While

  Return set of medoids C;

End

N and C are two neighbors, they differ only by one medoid, to get the set of medoids N we change in C the medoid c by the node n, thus the cost differential of the two neighbors N and Cis equal to the swapping cost of c by n (S_{cn})defined in the algorithm2, therefore Cost(C,N)=S_{cn} .

Raymond *et al.*, define in [12] the value of Maxneighbor like:

$$Maxneighbor = Max((1.25\% * k * (n - k)); 250) \qquad (10)$$

With n is the total number of the sensor nodes and k is the number of cluster in the network, k is defined above in the equation (8).

4.2. Steady State Phase

The steady state phase will start directly after the end of the set up phase. The non selected node gathers data and transmits them to the medoid node in the cluster. The medoid node (cluster head) forwards all aggregated data of the cluster to the sink, for that it creates and broadcasts a time division multiple access (TDMA) schedule in order to specify a time slot for each member node to transmit data to the cluster head. Thusby

using TDMA technique in the level of MAC layer we can avoid collisions in the cluster during the transmission of data.

The CH role can be rotated among the sensor nodes near to the medoid node if the level of the remaining energy of this former becomes critical allowing to save energy of control message overhead to create new clusters in every round. Therefore, only cluster head selection will be necessary in the set up phase.

5. Results and Simulation

We compared the results simulation of our protocol to the existing protocols: leach, leach-c and pegasis. The simulation is done on ns2 [13], there are 500 nodes arbitrary distributed on a sensing area of 1000m×1000m and the sink position is 75m from the nearest sensor node, the standard IEEE 802.15.4 is adopted for the communication between nodes. The efficiency of our new protocol is based on the average energy consumption of the nodes and the number of data received by the sink. The parameters used in the simulation are shown in the table1:

Table 1. Simulation Parameters

Parameters	Values
Node distribution	Random
Network size	1000m×1000m
Initial energy	2 Joules
Total number of sensor nodes	500
sink location	75m
Radio transmission range	20m
E_{elec}	50nJ/bit
E_{amp}	100pJ/bit/m^2
Critical remaining energy	0.01J

We assume the following properties about our network:

- The sink knows the location of sensor nodes.

- All sensor nodes are equal in resources (sensing, processing and initial energy level).

- The sensor nodes and the sink are stationary: static networks.

- Small number of anchor nodes is present in the network to determine the coordinates of the sensor nodes.

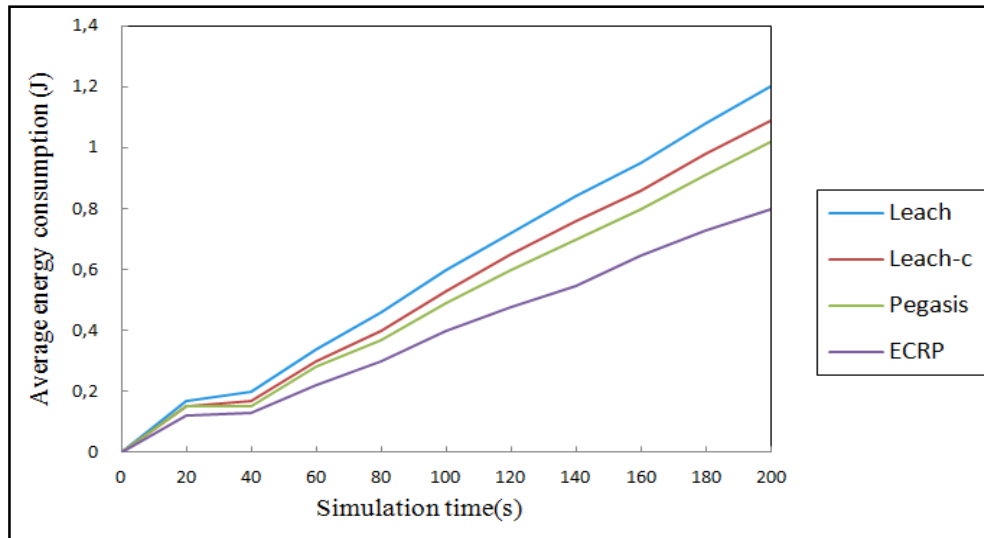- Sensor nodes collect and disseminate data with a periodic manner.

Figure 3. Comparison of Energy Consumption Through Time

The Figure 3 presents the average energy dissipation of sensor nodes through the time for the four protocols. It shows that our protocol optimizes the energy consumption better than the other protocols. Our protocol reduces the average distance between cluster head and the nodes in the cluster for that the operation of transmission of data becomes optimized and less expensive than the other protocols. This way, the energy dissipation of node to transmit data to the cluster head is reduced and optimized.
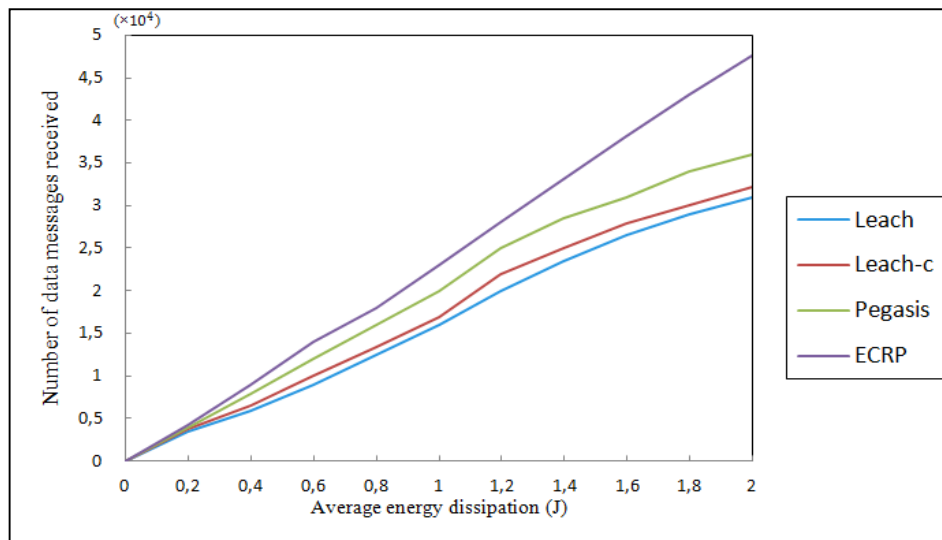

Figure 4. Quantity of Received Data by the Sink

We analyze the quantity of data message received by the sink. Figure 4 shows the number of data message received by the sink in function of the average energy dissipation for the four protocols. The results showed proven the efficiency of the protocol ECRP delivering more data to sink than the other protocols.

The two experimental results prove that our protocol outperforms the other hierarchical routing protocols and has the best performance.

6. Conclusion

Most of the clustering algorithms proposed in the literature are based on the election of a number of CHs amongst the sensor nodes of the network. With the proposed clustering method, the nodes of the same cluster transmit sensed data to their CH located in the center of the cluster allowing to minimize the data transmission cost to the cluster head. The latter aggregates the collected data by the sensors of its cluster in order to eliminate redundancy and sends it to the base station.

The ECRP protocol presented in this article outperforms the existing hierarchical routing protocol, we show that the use of a modified version of clarans algorithm in WSN achieves good results and distributes energy consumption evenly on the network.

References

[1] T. Ng. Raymond, and H. Jiawei, "Efficient and Effective Clustering Methods for Spatial Data Mining", Proceeding of International Conference on Very Large Data Bases, Santiago, Chile, **(1994)**.

[2] W. R. Heinzelman, A. Chandrakasan and H. Blalakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", Proceeding of the 33rd Annual Hawaii International Conference on System Sciences, Hawaii, USA **(2000)**.

[3] W. R. Heinzelman, A. Chandrakasan and H. Blalakrishnan, "An Application-Specific Protocol Architecture for Wireless Micro Sensor Networks", IEEE Transactions on Wireless Communication, Vol. 1, no. 4, **(2002)**, pp. 660-670.

[4] S. Lindsey and C. S. Raghavendra, "PEGASIS:Power-Efficient Gathering in sensor Information Systems", Proceeding of IEEE Aerospace Conference, Los Angelos, USA **(2002)**.

[5] S. Lindsey, C. S. Raghavendra and K. Sivalingam, "Data Gathering in Sensor Networks using the Energy *Delay Metric", in the Proceedings of the IPDPS Workshop on Issues in Wireless Networks and Mobile Computing, San Francisco, USA **(2001)**.

[6] O. Younis, S. Fahmy, "HEED: Hybrid Energy Efficient Distributed Clustering Approach for Ad hoc Sensor Networks", IEEE Transactions on Mobile Computing, vol. 3, no. 4, **(2004)**, pp. 366-379.

[7] D. Kumar, T. C. Aseri and R. B. Patel, "EEHC: Energy Efficient Heterogeneous Clustered scheme for wireless sensor networks", Computer Communications, vol. 32, no. 4, **(2009)**, pp. 662-667.

[8] G. S. Arumugam, T. Ponnuchamy, "EE-LEACH: development of energy-efficient LEACH Protocol for data Gathering in WSN", Eurasip Journal on Wireless Communication and Networking, **(2015)**, in press.

[9] A. Boukerche, H.A.B.F. Oliveira, E.F. Nakamura, and A.F.A. Loureiro, "Localization Systems for Wireless Sensor Networks", IEEE Wireless Communications, Vol. 14, no. 6, **(2007)**, pp. 6-12**.**

[10] Z. Fang, Z. Zhao, X. Cui, D. Geng, L. Du, and C. Pang, "Localization in Wireless Sensor Networks with Known Coordinate Database", EURASIP Journal on Wireless Communications and Networking, Vol. 2010, no. 2, **(2010)**.

[11] R. Misra, S. Shukla and V. Chandel, "Lightweight Localization Using Trilateration for Sensor Networks", International Journal of Wireless Information Networks, vol. 21, no. 2, **(2014)**, pp. 89-100.

[12] T. Ng. Raymond and H. Jiawei, "CLARANS: A Method for Clustering Objects for Spatial Data Mining", IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 5, **(2002)**, pp. 1003-1016.

[13] The Network Simulator – ns2. http://www.isi.edu/nsnam/ns/.

Authors

Rais Amine, he was born in Morocco in 1984, and got a master degree in information systems from Sidi Mohamed IbnAbdellah University – FEZ Morocco in 2008, he is now working toward the PHD degree in ENSEM – Casablanca Morocco, and his main interest fields include clustering and mobile agents in wirelesss sensor networks.

Bouragba Khalid, he is currently working as an assistant professor at the University Hassan II /ESTC, Casablanca Morocco. He received his PHD degree from University Hassan II/ENSEM – Casablanca Morocco. His research interests concern workflow systems and wireless sensor networks.

Ouzzif Mohammed, he is working as professor at the University Hassan II of Casablanca. He earned a PHD in the field of the specification and verification of collaborative systems at the University Mohamed V – Rabat Morocco. Currently, his research interests concern distributed systems and in particular the problems related to the wireless sensor networks in the context of smart cities.