

# POWER CONTROL AND CLUSTERING IN WIRELESS SENSOR NETWORKS

Lahcène Dehni<sup>1</sup>, Francine Krief<sup>2</sup>, Younès Bennani<sup>1</sup>

<sup>1</sup> *Laboratoire d'Informatique de l'Université Paris Nord, Institut Galilée, 99, Avenue J.B. Clément, 93430 Villetaneuse France ;* <sup>2</sup> *Laboratoire Bordelais de Recherche en Informatique Domaine Universitaire, 351, cours de la Libération, 33405 Talence Cedex, France*

**Abstract:** The use of the wireless sensor networks (WSNs) should be increasing in different fields. However, the sensor's size is an important limitation in term of energetic autonomy, and thus of lifetime because battery must be very small. This is the reason why, today, research mainly carries on the energy management in the WSNs, taking into account communications, essentially. In this context, we compare different clustering methods used in the WSNs, particularly EECS, with an adaptive routing algorithm that we named LEA2C. This algorithm is based on topological self-organizing maps. We obtain important gains in term of energy and thus of network lifetime.

**Key words:** Wireless sensor networks, clustering, power control, adaptive routing algorithm, and topological self-organizing maps.

## 1. INTRODUCTION

WSNs present a vast application field, for example, in the scientific, logistic, military and health field. According to MIT'S Technology Review<sup>1</sup>, *“this technology is one of the ten new technologies which will change the world and our manner of live and work”*.

The battery is an important component of a sensor. Generally, it is neither replaceable nor rechargeable. With its small size, it provides an energy quantity very limited on a scale of 2J by a sensor<sup>2</sup>. So, it limits the lifetime of the sensor and influences the total operation of the network. This is the reason why, today, protocols ensuring low energy consumption occupy an important research orientation in this field.

A sensor ensures acquisition, data processing and communications. The communications are the most energy consuming. Thus, a good diagram of energy management must, in priority, take into account communications.

The two principal classes of protocols used in the wireless networks are based on the multi-hops routing or clustering techniques. Several approaches are proposed to calculate the optimal path in the multi-hops routing protocols. Some<sup>3</sup> propose to take into account the shortest path, in term of distance, to the base station. Others<sup>4</sup> still, choose an optimal path by privileging some sensors whose presence in the path reduces the energy consumption.

However, the main disadvantage of the multi-hops routing is the periodicity of the messages sent to maintain the valid paths. These messages overload the network and consume additional energy.

Today, clustering gives the best results<sup>5</sup>; this is the reason why we adopted this approach. It is a problem of classification which interests the numerical learning for a long time, in particular, the connectionist models and, more particularly, the topological Self-Organizing Maps (SOMs). This last approach has proved its efficiency in this type of problem.

We proposed to adapt an unsupervised connectionist learning method by introducing the evolutionary and dynamic clustering aspect. This new approach of clustering allows us improving the efficiency of the routing in WSNs. In this paper, we compare the proposed routing algorithm with different clustering methods used in the WSNs, particularly EECS<sup>6</sup> (Energy Efficient Clustering Scheme in Wireless Sensor Networks). We obtain better results with our method.

Our paper is organized as follows: first, we present an energy consumption model in the WSNs. Then, we study different clustering protocols. We also present the unsupervised numerical learning technique for the clustering, called the self-organizing maps. The adaptation of this classification technique to the problem of routing allowed us to propose a routing algorithm adapted to the WSNs. Finally, we show, through an experiments series, some validation results of our new algorithm and we present the future prospects.

## 2. ENERGY CONSUMPTION MODEL IN WSNs

A sensor uses its energy to carry out three main actions: acquisition, communication and data processing. The power consumption to perform the data acquisition is not very important. Nevertheless, it varies according to the phenomenon observed and monitoring type. The communications consume much more energy than the other tasks, in emission as well as in reception. Fig. 1 presents an antenna model and the energy consumption rules associated<sup>9</sup>. To transmit a  $k$  bits message over a distance of  $d$  meters, the transmitter consumes:

$$E_{Tx}(k,d)=E_{Tx}(l)+E_{Tx\_amp}(k,d) \quad (01)$$

$$E_{Tx}(k,d)=\begin{cases} k.E_{elec}(k,d)+k.E_{friss}.d^2 & \text{if } d < d_{crossover} \\ k.E_{elec}(k,d)+k.E_{two\_ray\_amp}.d^4 & \text{else} \end{cases} \quad (02)$$

To receive a  $k$  bits message, the receiver consumes:

$$E_{Rx}(k)=E_{Rx\_elec}(k)=k.E_{elec} \quad (03)$$

Where :  $E_{elec}$  : energy of electronic transmission/reception;  $k$  : size of a message;  $d$  : distance between the transmitter and the receiver;  $E_{Tx\_amp}$  : amplification energy;  $\epsilon$  : amplification factor;  $d_{crossover}$  : limit distance over which the transmission factors change of value.

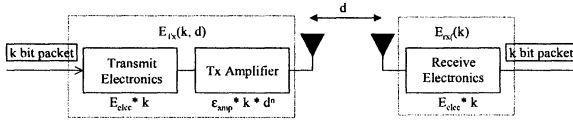


Figure 1. Communication power consumption model in WSNs

Power consumption in aggregation is calculated by:  $E_{DA} = 5 \text{ nJ/bit/msg}$  (04)

### 3. CLUSTERING PROTOCOLS USED IN WSNs

Heinzelman and al.<sup>8-9</sup> have proposed the LEACH protocol and a centralized version of this protocol, called LEACH-C. These protocols are based on clustering.

Clustering consists in the segmentation of the network into groups (clusters). Sensors transmit their data towards group representatives called clusterheads (CHs), which send these data to the base station (BS). In some applications, CHs make a simple data processing (aggregations...) on the received data before retransmitting them to the BS. Clustering permits the bandwidth re-utilization. It also offers a better resource allocation and helps to improve the energy control in the network<sup>10-11</sup>. With the clustering sensors establish small communication distances with their CHs. The data aggregations and compressions minimize energy consumption by reducing the data flow and thus the total communications.

LEACH protocol is based on a probabilistic demand model and fixes the optimal clusters number according to some parameters such as the network topology, the communications and the computational cost. (Generally, CHs represent 5% of the sensors number in the network). CHs create TDMA tables according to the number of sensors in a cluster. Each sensor transmits data to CH using the time slots specified in the TDMA tables. Sensors turn off their antennas and wait for their speaking time. It permits to minimize the energy dissipation. CHs leave their receivers on to receive the sensors data.

Each CH chooses randomly a code in a list of CDMA propagation codes; it transmits this list to its sensors in the cluster. The sensors use this list for their transmissions. It permits to minimize the communication interferences between closed CHs. The use of TDMA/CDMA techniques allows a hierarchy built on a multilevel clustering, it can increase the amount of saved energy.

An extension of this algorithm, called LEACH-C<sup>9</sup>, has been proposed to avoid these drawbacks. In this centralized iterative algorithm, the clusters structure

is computed by the BS using the "Simulated annealing" optimization method<sup>12</sup>. At each step, the sensors are given their role, either CH or simple sensor, by the BS. Then, operations continue like in the LEACH protocol.

Authors<sup>6</sup> present EECS (Energy Efficient Clustering Scheme in WSNs). It is a LEACH-like clustering scheme. CHs are elected depending on their communication residual energy while achieving a well CH distribution; furthermore, a new method is introduced in order to balance the load among the CHs. The simple sensors choose their CH by considering not only saving its own energy but also balancing the workload of CHs. To do that, two distance factors are introduced:  $d(P_j, CH_i)$  (distance between a node  $P_j$  and a  $CH_i$ ) and  $d(CH_i, BS)$  (distance between a  $CH_i$  and the BS) and two proposed normalized functions to get a *cost* function.  $P_j$  chooses  $CH_i$  with  $\min \{cost\}$ . Simulation results show that EECS outperforms LEACH significantly and prolongs the network lifetime over 35%<sup>6</sup>.

#### 4. THE UNSUPERVISED CONNECTIONIST LEARNING AND THE SOM

Unsupervised numerical learning, or automatic classification, consists in determining a partition of an instance space from a given observations set, called training set. It aims to identify potential trend of data to be gathered into classes. This learning approach, namely *clustering*, seeks for regularities from a sample set without being driven by the use of the discovered knowledge. Euclidian distance is usually used by clustering algorithms to measure similarities between observations. Self-Organizing Maps (SOM) implement a particular form of competitive artificial neural networks; when an observation is recognized, activation of an output cell – competition layer – leads to inhibit activation of other neurons and reinforce itself. It is said that it follows the so called "Winner Takes All" rule. Actually, neurons are specialized in the recognition of one kind of observations.

The learning is unsupervised because neither the classes nor their number is fixed a priori.

A SOM consists in a two dimensional layer of neurons (Fig. 2) which are connected to  $n$  inputs according  $n$  exciting connections of respective weights  $w$  and to their neighbors with inhibiting links.

The training set is used to organize these maps under topological constraints of the input space. Thus, a mapping between the input space and the network space is constructed; closed observations in the input space would activate two closed units of the SOM.

An optimal spatial organization is determined by the ANN from the received information, and when the dimension of the input space is lower than three, both position of weights vectors and direct neighborhood relations between

cells can be represented visually. Thus, a visual inspection of the map provides qualitative information of the map and the choice of its architecture. The connectionist learning is often presented as a minimization of a risk function. In our case, it will be carried out by the minimization of the distance between the input samples and the map prototypes (referents), weighted by a neighborhood function  $h_{ij}$ . To do that, we use a gradient algorithm. The criterion to be minimized is defined by:

$$E_{SOM} = \frac{1}{N} \sum_{k=1}^N \sum_{j=1}^M h_{j, NN(x^{(k)})} \|w_{.j} - x^{(k)}\|^2 \quad (06)$$

$N$  represents the learning samples number,  $M$  the neurons number,  $NN(x^{(k)})$  is the neuron having the closest referent to the input form  $x^{(k)}$ , and  $h$  the neighborhood function. The neighborhood function  $h$  can be defined as:

$$h_{r,s} = \frac{1}{\lambda(t)} \exp\left(-\frac{d_1^2(r,s)}{\lambda^2(t)}\right) \quad (07)$$

$\lambda(t)$  is the temperature function modeling the neighborhood extent, defined as:

$$\lambda(t) = \lambda_i \left(\frac{\lambda_f}{\lambda_i}\right)^{\frac{t}{t_{max}}} \quad (08)$$

$\lambda_i$  and  $\lambda_f$  are respectively initial and the final temperature.  $t_{max}$  is the maximum number allotted to the time (number of iterations for the  $x$  learning sample).  $d_1(r,s)$  is the Manhattan distance defined between two neurons  $r$  and  $s$  on the map grid, with the coordinates  $(k,m)$  and  $(i,j)$  respectively:

$$d_1(r,s) = |i-k| + |j-m| \quad (09)$$

The learning algorithm of this model proceeds essentially in three phases:  
 - Initialization phase where random values are assigned to the connections weights of each neuron of the map grid.  
 - Competition phase during which, for any input form  $x^{(k)}$ , a neuron  $NN(x^{(k)})$ , with neighborhood  $V_{NN(x^{(k)})}$ , is selected like a winner. This neuron has the nearest weight vector by using Euclidean distance:

$$NN(x^{(k)}) = \underset{1 \leq i \leq M}{\operatorname{argmin}} \|w_{.i} - x^{(k)}\|^2 \quad (10)$$

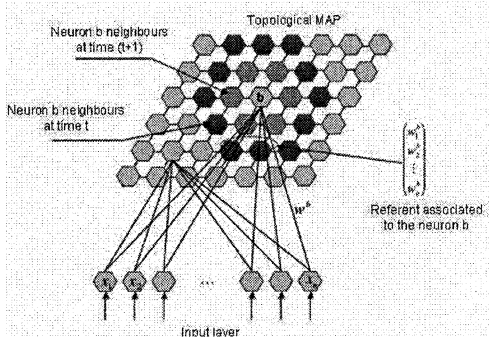


Figure 2. MAP structure (2D network)

- Adaptation phase where the weights of all the neurons are updated according to the following adaptation rules: If  $w_{.j} \in V_{NN(x^{(k)})}$  then adjust the weights

$$\text{using: } w_{.j}(t+1) = w_{.j}(t) - \varepsilon(t) h_{j_{NN(x^{(k)})}}(w_{.j}(t) - x^{(k)}) \quad (11)$$

$$\text{else } w_{.j}(t+1) = w_{.j}(t) \quad (12)$$

Repeat this adjustment until the SOM stabilization.

- **SOM map segmentation:**

We segment the SOM using the  $K$ -means method (Fig. 3). It is another clustering method. It consists in choosing arbitrarily a partition. Then, the samples are treated one by one. If one of them becomes closer to the center of another class, it is moved into this new class. We calculate the centers of new classes and we reallocate the samples to the partitions. We repeat this procedure until having a stable partition.

The criterion to be minimized in this case is defined by:

$$E_{K\text{-means}} = \frac{1}{C} \sum_{k=1}^C \sum_{x \in Q_k} \|x - c_k\|^2 \quad (13)$$

Where  $C$  represents the number of clusters,  $Q_k$  is the cluster  $k$ ,  $c_k$  is the center of the cluster  $Q_k$  or the referent.

The basic algorithm requires fixing  $K$ , the number of clusters wished. However, there is an algorithm to calculate the best value for  $K$  assuring an optimal clustering. It is based principally on the minimization of Davies-Bouldin index, defined as follows:

$$I_{DB} = \frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad (14)$$

$$\text{With } S_c(Q_k) = \frac{\sum_i \|x_i - c_k\|^2}{|Q_k|} \quad (15)$$

$$d_{cl}(Q_k, Q_l) = \|c_k - c_l\|^2 \quad (16)$$

$C$  is the number of clusters,  $S_c$  is the intra-cluster dispersion, and  $d_{ce}$  is the distance (centroid linkage) between the clusters centers  $k$  and  $l$ .

This clustering procedure aims to find internally compact spherical clusters which are widely separated. There are several methods to segment the SOMs<sup>13</sup>. Usually, they are based on the visual observations and the manual assignment of the map cells to the clusters. Several methods use the  $K$ -means algorithm with given ranges for  $K$  value. Our work is based on the approach of Davies-Bouldin index minimization<sup>14</sup>.

We note that the  $K$ -means approach can be applied directly to the data instead of SOMs. In our work, we applied it to the SOMs results. The idea is to use SOMs as a preliminary phase in order to set a sort of data pretreatment

(dimension reduction, regrouping, visualization...). This last has the advantage to reduce the clusters calculation complexity and also ensure a better visualization of the automatic classification results. Moreover, the use of SOMs for visualization is crucial, especially in the case of data multivariate: dimension  $> 2$  or 3. In this last case, the SOMs permit to reduce the data space dimension and to visualize the clusters in the plan.

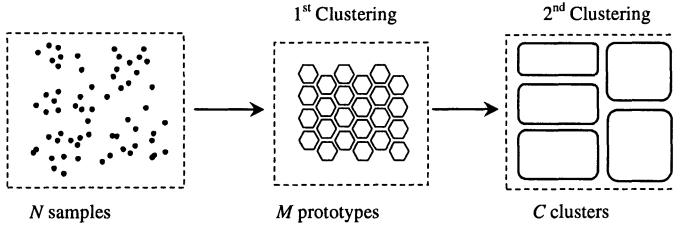


Figure 3. Two successive clusterings: SOM followed by K-means

## 5. LEA2C: A NEW ROUTING APPROACH IN WSNS

We use clustering methods based on the unsupervised connectionist learning techniques and different properties of LEACH-C in order to propose a new routing approach in WSNs, so called LEA2C (Low Energy Adaptive Connectionist Clustering). In our approach the sensors have a GPS system (Global Positioning System). It allows the BS to localize the sensors. As in LEACH-C protocols<sup>8-9</sup>, we use the power consumption model of the Fig. 1.

Our algorithm is iterative: In each iteration the BS calculates the clustering according to sensors alive coordinates. It affects roles to the sensors by assigning transmission codes and frequencies. Sensors clustering is done with the SOMs and the number of clusters is optimized using the  $k$ -means algorithm. In each cluster, the CH choice can be made using one of the three following criteria: - The sensor having the maximum power level in the cluster. - The nearest sensor to the gravity center of the cluster or to the BS.

*LEA2C algorithm:*

To compare our approach with other methods such as EECS method, we use the same communication and data model; also we run our simulation in two different scenes<sup>6</sup>.

**1) Initialization:** Random deployment of the  $N$  homogeneous sensors in a given space and with the same energy level.

**2) Clustering**

**2.1)** Clustering of the WSN by using the SOM and  $K$ -means algorithms by using the sensors coordinates.

**2.2)** CHs selection according to one of the selection criteria cited above.

**2.4)** Roles affecting to each the sensors (CH or simple sensor).

**3) Data transmission**

- 3.1) Data transmission from the simple sensors to the CHs: consumed energy is calculated using the Eqs. (01) and (02).
- 3.2) We calculate, for each CH:
  - The data reception and aggregation energy using the Eq.(03, 04)
  - The results transmission energy to the BS by using the Eq.(01).
- 3.3) When the CH is chosen according to the first criterion (maximum energy), the same CHs are reelected after each transmission.
- 3.4) Repeat the steps 3.1) to 3.4) until the death of one sensor.
- 4) Repeat the steps 2) and 3) until the death of all the sensors in the network.

### 6. SIMULATIONS RESULTS

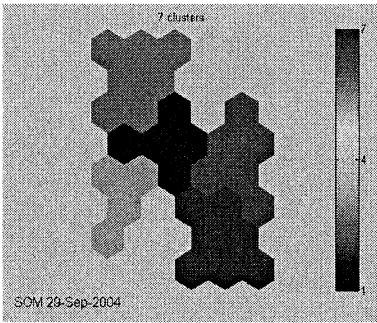


Figure 4.a. SOM clustering

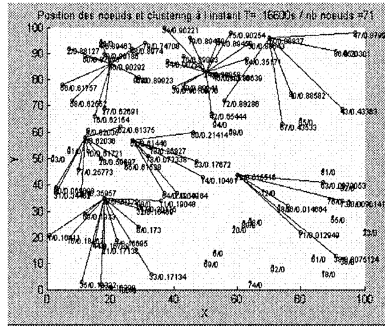


Figure 4.b. LEA2C communications

To simulate our algorithms, we have used SOMs core (SomToolbox) proposed by HUT researchers<sup>15</sup> (Helsinki University of Technology). We have applied our algorithm on the same data used in EECS<sup>6</sup>. The figures below present the results obtained.

Fig. 4.a. represents the neurons segmentation of the SOM map.

Fig. 4.b. represents the sensors network, and shows the sensors clustering.

Figs. 5.a and 5.b represent respectively the variation of the number of sensors alive according to the number of received messages by the BS in the normal scene (400) and the large scene (1000) motes. In each figure, we consider the corresponding variations according to the four scenarios:

- Direct transmission from the sensors to the BS.
- The clustering according to our approach with the three CH selection criteria already mentioned.

Figs. 6.a and 6.b represent the same variations with the EECS protocol.

In general, Figs. 5.a and 5.b show, for our algorithm, that the choice of the CH according to the criterion of maximum energy is better than the choices according to the other criteria. Choices according to the criteria of proximity



to the GC and to the BS have nearly the same performances. They are enough far from the criterion of maximum energy but better than the performances of the direct transmission protocol and EECS ones.

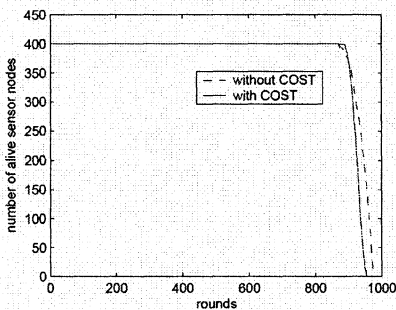


Figure 6.a Number of the alive nodes VS time with EECS (First scene)<sup>8</sup>

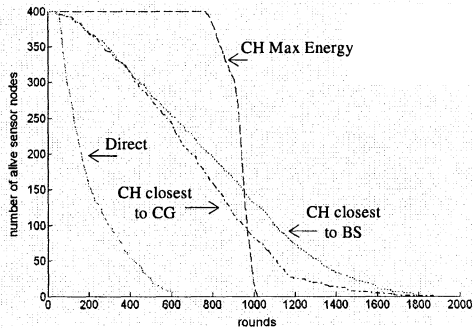


Figure 5.a Number of the alive nodes VS time with LEA2C (First scene)

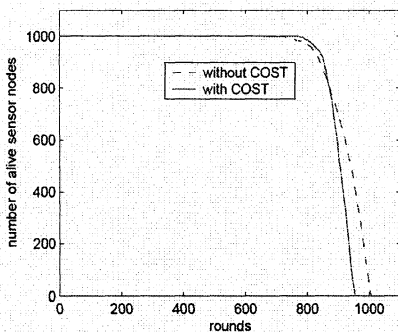


Figure 6.b Number of the alive nodes VS time with EECS (Second scene)<sup>8</sup>

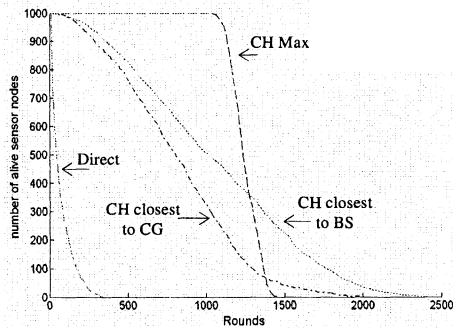


Figure 5.b Number of the alive nodes VS time with LEA2C (Second scene)

Figs. 6.a and 6.b show the effectiveness of the EECS. Moreover, EECS with *cost* gives better results than the *without-cost* scenario.

By comparing the Figs. (5.a, 5.b) and Figs. (6.a, 6.b), we notice that the graphs have the same paces with very apparent profits brought by our algorithm compared with the EECS protocol.

On the graphs 6.a and 6.b, we notice that LEA2C, with the better selection criterion of CHs (Max energy level), makes it possible to ensure a totality survival of sensors during 90% of the network lifetime. Compared to EECS, our algorithm offers a profit of 50% more important in term of total network lifetime when the CH retained is with the max energy level, and 100% times more when the CH retained is the nearest sensor to the BS.

## 7. CONCLUSION AND PROSPECTS

LEA2C ensures an important profit compared with the EECS protocol. The results obtained are very promising with a lifetime up to 50% longer. The network coverage is insured during 90% the total time of treatment.

To improve our results, we plan the following adaptations: - Application of a super-clustering on the CHs, and the spreading of the clustering over several levels. - Use of other SOMs versions and the coupling with other communication protocols. - Optimization of the parameters of the learning algorithms (SOMs,  $K$ -means) - Integration of other parameters in the clustering process, such as the moving speed of the sensors in the case mobiles sensors.

## 8. REFERENCES

1. Technology Review: 10 Emerging technologies that will change the world (February 2003); <http://www.technologyreview.com>.
2. G.J. Pottie and al. Wireless integrated network sensors; *Communications of the ACM* 43 (5), pp. 551– 558. (2000)
3. C. Perkins and E. Royer, Ad-Hoc on-demand distance vector (AODV) routing; *The Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99)*. (1999)
4. K. Scott and N. Bambos, Routing and channel assignment for low power transmission in PCS; *5th IEEE Int. Conf. on Universal Personal Communications*, volume 2. (1996)
5. S. Ghiasi et al. Optimal energy aware clustering in sensor networks; *SENSORS Journal*, Vol. 2, Issue 7, pp. 258-269, July 2002.
6. M. Ye, C. Li, G. Chen and J. Wu, EECS: An energy efficient clustering scheme in wireless sensor networks; *IEEE IWSEASN'05*. (2005)
7. I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, A survey on sensor networks"; *IEEE Communications Magazine*, Vol. 40, No. 8, pp. 102-114, (2002).
8. W. Heinzelman, A.P. Chandrakasan and H. Balakrishnan; Energy-efficient communication protocol for wireless microsensor networks; *Sensor 2002*, 2, pp. 258 –269. (2002)
9. W. Heinzelman, A.P. Chandrakasan and H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks; *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 660-670. (2002)
10. E.-S. Jung and N. H. Vaidya, A power control MAC protocol for ad-hoc networks; *ACM MOBICOM*. (2002)
11. V. Kawadia and P. R. Kumar, Power control and clustering in Ad Hoc networks; *IEEE INFOCOM*. (2003)
12. T. Murata and H. Ishibuchi, Performance evaluation of genetic algorithms for flowshop scheduling problems; *1st IEEE Conference Evolutionary Computation*, volume 2. (1994)
13. A. Juha and A. Esa, Clustering of the self-organizing map; *IEEE Transactions On Neural Networks*, volume 11, n° 3, (2000)
14. David L. Davies and Donald W. Bouldin, A cluster separation measure; *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-1(2): pp. 224-227. (1979)
15. E. Alhoniemi and al. SOM Toolbox, (2000). <http://www.cis.hut.fi/projects/somtoolbox/>