2011 International Conference on Network Computing and Information Security

General criteria-based clustering method for sensor network

Yu Niu, Brian J. d'Auriol, Sungyoung Lee
Dept. of Computer Engineering
Kyung Hee University
Suwon, South Korea
{*niuyu, dauriol, sylee*}*@oslab.khu.ac.kr*

***Abstract*—This paper proposes a General-Criteria based Clustering(GCC) method for wireless sensor networks by intimating the activities of biological neurons. The GCC expands the option of clustering criteria and the result cluster configuration. The generality of criteria enables the application on-demand clustering become possible. Based on that a novel cluster concept *Logical Cluster* is proposed. The simulation shows that the GCC can get diverse logical clusters and synchronization/desynchronization (sync/desync) coexistence results with acceptable energy consumption.**

***Keywords*-clustering; synchronization/desynchronization; sensor network;**

I. INTRODUCTION

Node clustering, which aims to increase scalability and reduce the complexity of network management, is common in sensor networks. Most of the existing clustering algorithms divide the network nodes by applying some simple criteria, such as node location and communication costs.

However, the often used criteria ignore the nature and characteristics of the sensor nodes as well as the requirements of the application level. Moreover, many existing techniques require additional assistant algorithms for communication in cluster networks. For example, some need a central node for controlling or special coding methods to avoid the collision among the nodes and clusters. Some others need additional time synchronization algorithms for data aggregation or other data recording functions.

In this paper, we propose a general criteria-based clustering (GCC) method. In this method, the neuron oscillators' interactive coupling activity is imitated to produce independent clusters based on the oscillators' initial phase distribution. Because the phase is an abstract concept for a mathematical system, its initial value can be mapped from any sensor node related data or properties, which means any node related characteristic data can work as the clustering criteria rather than just the node geography information. The proposed method is general criteria-based and the GCC method overcomes the limitations in the existing clustering techniques by considering the nature and characteristics of sensor nodes. In particular, application level requirements can be used to influence the results of clustering. The use of the application level requirements leads to a connection between applications and clustering. On demand and dynamic changes in the applications may also be reflected in re-clustering.

Besides the clustering, the GCC method has another advantage. Within a cluster, the members are phase synchronized, and among clusters, each cluster is phase desynchronized. If needed, these synchronized/desynchronized phases can also be reversely mapped as the sensor nodes' logical clock. This enables the applications that need time synchronization or desynchronization without requiring any additional algorithms or procedures.

The main contributions of this paper include: 1) propose a new GCC method and the concept of Logical Cluster (LC); 2) achieve local synchrony and global de-synchrony in a network at the same time; 3) for the first time, the delayed inhibitory pulse coupling oscillator model in the neuron system is introduced to the field of sensor networks.

The following paper is set as next, section II introduces the existing clustering algorithms and neuron model's application in sensor networks as prior work. The inhibitory pulse coupling model and clustering results are detailed in section III. Section IV lists the GCC method operations step by step. Section V focuses on the analysis and simulation of the proposed GCC method and section VI summaries the work and discusses some of the future work.

II. PRIOR WORK

In traditional clustering algorithms, the node divisions and cluster formations are led by the cluster head. Nodes which have closer distances to the designated head node are grouped together. The only criterion for deciding which cluster a node belongs to is its distance to the head node or the communication cost. The main work in traditional clustering algorithms is to choose proper cluster head. Some algorithms choose cluster head based on the residual energy, some are based on node degree and some are based on the combine weights of several critical factors. All of these algorithms partition the network area geometrically and the correlation between cluster formations and the sensing environment or the sensing data are not considered.

Recently, the mutual interaction between neurons inspired pulse coupling oscillator model are studied and introduced in the field of sensor networks. However, most of the work

978-0-7695-4355-0/11 $26.00 © 2011 IEEE
DOI 10.1109/NCIS.2011.153

focus the applications on time synchronization. [1] proves that through very simple reactive adjustments of the node phases after receiving the firing pulse, the phases of all oscillators would converge to a global synchronicity, regardless of the number of nodes and their initial states. Our previous work [2] gets rid of the swing actions in the coupling procedure by predicting the final converging direction and improves the synchronizing speed and energy efficiency. To eliminate the ideal assumption of instantaneously coupling, nodes in [3] accumulate the incoming pulses in the past period and do the phase adjustment once in all at the begin point of the next period. [4] enlarges the period from 1T to 2T, when the node gets a pulse at some point of the first T, it will react to the pulse at the same point of the second T. The extended T is used to buffer the transmission delay.

All of the above-mentioned work use the same kind of pulse coupling model: monotonically increase, concave down function figure, instant coupling and excitatory phase adjustment. However, all these components can have alternative choices and there are many variations in the dynamic systems.

III. Delayed Inhibitory Pulse coupling Model

In the pulse coupling model, every sensor node is regarded as an oscillator whose potential state $x_i(t)$ increases in $[0,1]$ along with its phase $\phi_i(t)$ periodically. When the potential $x_i(t)$ reaches the threshold x_{th}, the oscillator emits a fire pulse and falls back to 0 and restarts again. Figure 1 shows the curve of an oscillator potential (Y axis) along with its phase (X axis). The potential is written a function of the phase ϕ:

$$x_i(t) = f(\phi_i(t)) \tag{1}$$

In which the phase is defined in the range of $[0,1]$ and the function f curve is monotonically increase and concave down ($f' > 0$, $f'' < 0$). Here, set $x_{th} = 1$ so $f(\phi) \in [0,1]$ and $f(0) = 0$, $f(1) = 1$.

When isolated, the oscillator repeats the shift, fire, reset activities round by round. But if it is in a network, when oscillator i fires at time t, all its neighboring nodes will react to the fire signal after some delay τ. This delay is critical for producing the clustered result, because without the delay, the instantaneous coupling will cause system converge to synchrony or chaotic asynchrony state. If neighbor oscillator $j's$ phase is $\phi_j(t')$ at time $t' = t + \tau$, after the coupling its phase will adjust as:

$$\phi_j(t'^{+}) = \begin{cases} f^{-1}(f(\phi_j(t')) + \frac{n_0(t)}{n}\epsilon) = B, & \text{if } 0 < B < 1 \\ 0, & \text{otherwise} \end{cases}$$

Because the phase is bounded in $[0,1]$, if B's value is less than 0 or bigger than 1, it should be regarded as 0. n is the network node number. Factor $0 < \frac{n_0(t)}{n} < 1$ controls dynamic coupling strength according to the simultaneously fired node number $n_0(t)$ at time t. The sign of ϵ is another key factor to influence the converging results. When $\epsilon > 0$, the potential will jump up, the phase jumps forward and the integrating period will be shortened. The coupling is excitatory. With excitatory coupling, the final converged cluster state is unstable. When $\epsilon < 0$, the potential jumps down, the phase jumps backward and the period is elongated. The coupling is inhibitory (figure 1) and the final clustered state is stable.

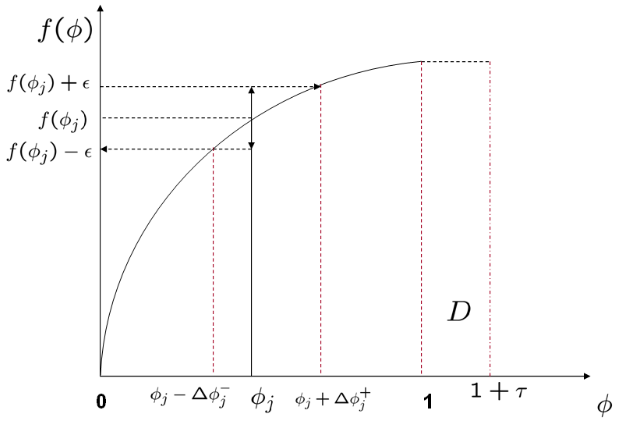


Figure 1. Potential $f(\phi)$ and phase ϕ adjustment during excitatory/inhibitory coupling.

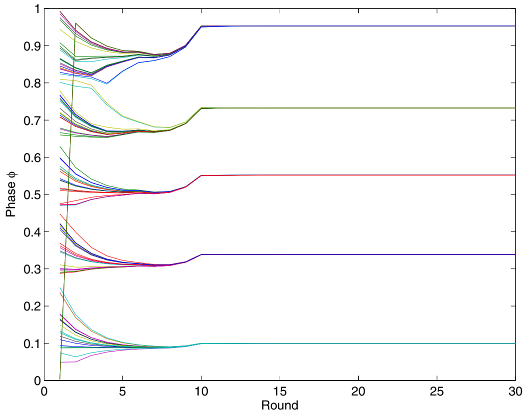


Figure 2. The converging procedure and clustered result under delayed inhibitive coupling.

Reference [5] observed that when the coupling is inhibitory and always has τ delay after firing, the oscillators in the whole network will converge to several clusters. All members in one cluster share the same phase. Between clusters, each pair of clusters take the fixed phase offset. They are phase desynchronized (figure 2). And this sync/desync state is stable because either the outside noises or minor phase deviations can not break down the clustered state. Besides, empirical data show that the number of converged cluster m is approximately inversely proportional to twice the length of delay τ:

$$m \approx 1/2\tau \tag{2}$$

IV. GENERAL CRITERIA-BASED CLUSTERING (GCC) METHOD

From the above discussion and figure 2 we can see the nodes are are clustered based on their initial phases and the final clustering results are controlled by several system parameters. This section gives the detailed steps of the method.

Step 1 Clustering Criteria Selection

Like stated above, GCC builds a connection between application requirements and clustering construction because of the general criteria. Because any node related data can work as the network clustering criteria. User or system can select clustering criteria based on the application level requirements and the expected benefits. Here gives out some examples that show how the different criteria bring different clustering results and benefits.

1) Selecting geographical information works as the clustering criteria. The geography information not only means the coordinates of the nodes' x, y but also includes the altitude, the distance to some point or area, or angle to some base line, etc. And the outcoming clusters may have the polygon shape as the usual clusters or some other various shapes. This kind of clustering could used for the surveillance on some hot spots, like fire event data gathering or traffic jam evacuating.

2) Selecting the node residual energy as clustering criteria. The nodes are grouped based on their energy level. Cluster members can adjust the sampling and reporting rate according to their energy level so as to balance the workload and extend the network lifetime.

3) Selecting the sensed content data as the clustering criteria. It will be easier to do data aggregation inside clusters whose members have similar content data. The final condensed information would be shorter and cost less transmission energy.

All the other numerable characteristics about the sensor nodes can also be the criteria of clustering, like the node id, their degree, etc. But whatever the criteria property is selected, the members in one cluster have close value of this property. Although the member nodes may not physically locate close as that in the normal cluster, they are logically related. We call this new kind of cluster as Logical Cluster (LC).

Step 2 Mapping and Parameter Setting

After selecting the appropriate clustering criteria, next step is to map the real criteria data as the initial phase values and set parameters in both physical and mathematic system.

When mapping the initial phase value ϕ, the network should be aware of the possible minimum and maximum values of the criterion as boundaries. The values in $[min, max]$ are normalized into $[0,1]$. These boundary values are calculated prior to deployment or obtained from the broadcasted message before clustering starts. Some node property data may be more than one dimension (like X,Y-coordinates) or the network selects more than one property as the criteria. Although there are some mathematical work on multi-dimension phase convergence, it is still too complicated. Therefore, the dimension decreasing is required.

After getting the initial phases, the phase moving speed is needed to calculate consequent phases value. The ordinary moving speed of the phase is decided by the oscillating period T and $\frac{d\phi_i}{dt} = 1/T$. The oscillating period T takes the same value T_0 in both the physical system and the mathematic model. The vale of T_0 can be set rather freely under the constraint of the system delay and the expected cluster number m. Also the length of T_0 will influence the whole system converging time and the corresponded energy consumption, which will be discussed in the following paragraphs.

Delay are inevitable in most real networks, especially in wireless networks. The GCC method considers the transmission delay directly in basic coupling model, in which after some nodes fire, instead of reacting spontaneously, receivers wait some time and then react. In real network, transmission delay mainly includes: transmitting delay t_{tx} in the sender, propagation delay t_{pg} in the air and decoding delay t_{dec} in the receiver. Compared to the t_{tx} and t_{dec}, the propagation time is small enough to neglect. On the other hand, t_{dec} is measured by the receiver itself and the length is known. The only nondeterminate item is the t_{tx}. If we want to get the exact t_{tx} value for every signal, the exact timestamps can be piggyback in the signal message. However, in the GCC method, all the fire messages are the same, being the short, pulse-like signals. Their emitting and receiving can be done by hardware or firmware using almost the same time. If the precision requirement is not very high, for simplicity, the transmission delay of all nodes are considered the same. $t_0 = t_{tx} + t_{dec}$ represents the total delay for a firing signal between its emission and reception. t_0 is the minimum allowed delay length and the real delay value t_τ shouldn't be shorter than it: $t_\tau \geq t_0$. Normalize the time delay for system model:

$$\tau = \frac{t_\tau}{T_0} \tag{3}$$

m is the expected cluster number. Considering $\tau \in [0, 0.5]$ and the relation $m \approx 1/2\tau$, the delay value also sets the floor boundary of the real oscillating period T_0.

$$T_0 \approx 2mt_\tau \geq 2mt_0 \tag{4}$$

Table I shows the mapping relationship between the physical parameters and system parameters.

Step 3 Pulse Coupling

To guarantee the convergence of the oscillators to the clusters, the system requires that the coupling is the all-to-all form, which means when one node fires and emits a signal, all nodes in the network should receive and react to

Table I
MAPPING FROM PHYSICAL TO SYSTEM PARAMETER

Physical parameter	**System parameter**
Criteria property $[min, max]$	Phase ϕ $[0, 1]$
Oscillating period T_0	T_0, $\frac{d\phi}{dt} = \frac{1}{T_0}$
$t_{delay} = t_{tx} + t_{pg} + t_{dec}$ $\approx t_{tx} + t_{dec} = t_0$ $t_\tau \geq t_0$	$\tau = \frac{t_\tau}{T_0}$ $\in [0, 0.5]$
Cluster number m	$m \approx \frac{1}{2\tau}$

it. There may emit lots of signals in the channel. So the firing signal should be short and fast travel in long distance.

One of the options is Ultra WideBand (UWB) pulse radio. UWB uses large bandwidth (>500MHz) short duration pulses to transmit information and its high data rate can be traded for range by aggregating pulse energy per data bit. This means besides high data rate in short range UWB can still work for longer range at a low data rate. In GCC, the firing signals do not need contain any source or target information that they just inform other nodes its firing time point. So the long transmission range and low bit rate UWB pulse radio meets our requirements of the firing signal. However taking UWB radio in sensor network needs extra software and hardware support which is not standard configuration for normal nodes and this will increase the total cost.

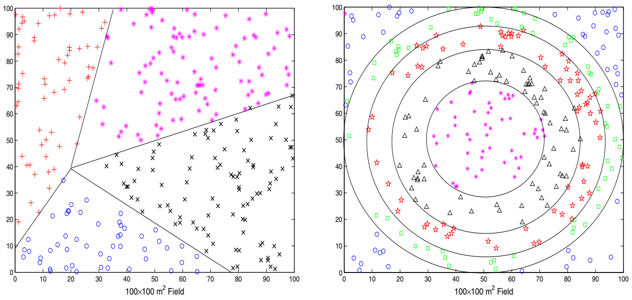
The other way is to use some specific sequences of pulses as the firing signal. The IEEE 802.11g protocol propose two synchronization head formats. One is the long version with 128 bits and the short one has 56 bits. Except the arranged bit sequence, the head file doesn't contain any other information. It is low cost and supported by existing popular transmission protocol. Therefore, we choose the short preamble as our firing signal in our following simulation.

Step 4 Reverse Mapping

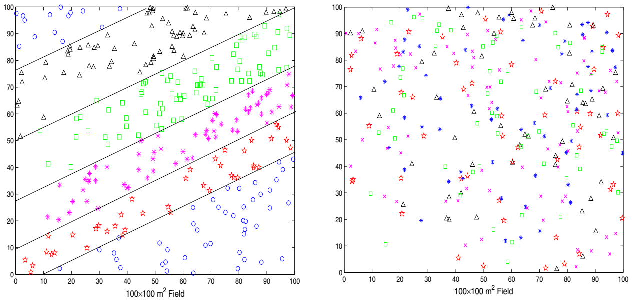
As mentioned in the introduction part, the members are phase synchronized inside a cluster. If the application requires time synchronization among nodes, the phase can be reversely mapped to some time variables. Because the member nodes' phase and frequency are all the same, started from a common zero point, the transformed time scale must be the same too. All nodes are time synchronized.

V. SIMULATION RESULTS AND DISCUSSION

The simulation environment is set as follows, 300 nodes are randomly deployed in a $100 \times 100m^2$ filed. The field is set as a single cell where firing signals are directly broadcasted and received and no relay work in it. Each node uses the uniform oscillating function $f(\phi) = \frac{1}{b}ln[(e^b - 1)\phi + 1]$, in which $b = 3$ measures the extent to which f is concave down. And the firing signal is chosen to be the 802.11g 56 bits short preamble.



(a) Angle to (20,40) and y=40. (b) Distance to point(50,50)



(c) Distance to line y=1/2x+20 (d) Node ID, residual energy...

Figure 3. Various shaped logical clusters based on different criteria

A. General-Criteria Clustering

To check the generality of GCC's criteria, here we choose several representative criteria as examples.

The X and Y axes in each subfigure of figure 3 are the field coordinates. In figure 3(a), the geography data criteria is the angle to the base line of a polar coordinate (defined by point $(20, 40)$ and line $y = 40$). The produced polygon clusters have no difference with the normal ones except for no predesignated head nodes. This example shows that the GCC method is compatible with the traditional clustering algorithms. Figure 3(b) and 3(c) choose the distance to field point $(50, 50)$ and line $y = 1/2x + 200$ respectively as the criteria. The formed cluster shapes are concentric circles and parallel strips respectively. Inside the cluster, head centric star topology may not be suitable any more. Instead a chain liked topology is more practical for strip shaped clusters. The duty of head node can shift along the chain for each round. Figure 3(d) shows the key idea of logical cluster. The clustering criteria can be the node ID, residual energy or other location unrelated properties. Physically the cluster members scatter everywhere. Logically, they have close criteria property. The overlapped physical range makes the clusters can not be active at the same time. The alternate sleeping schedule between clusters can solve this problem and save energy meantime.

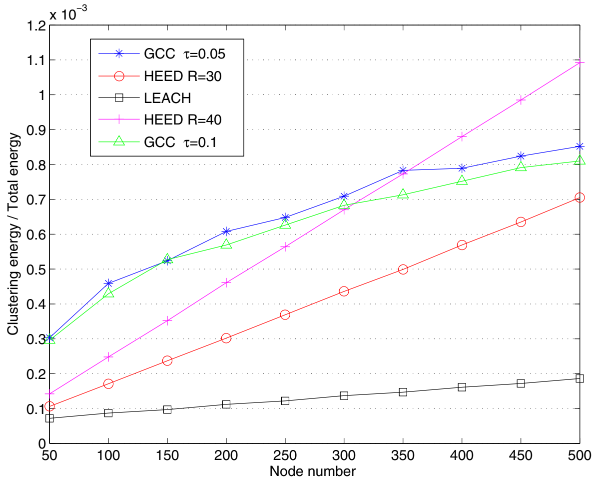


Figure 4. Energy cost ratio comparison under different node density. (GCC $\epsilon = 0.4$; LEACH 5% nodes as head)

B. *Energy consumption*

Next we compare the energy consumption situation of the GCC method and the traditional clustering algorithms. Two representative distributive clustering algorithms HEED (Hybrid, Energy-Efficient, Distributed Clustering) Approach and LEACH (Low Energy Adaptive Clustering Hierarchy) protocol in sensor networks are selected for comparison purposes. For HEED the fixed power level is used for intracluster communication and choose min-degree as the secondary communication cost. For HEED and LEACH, the broadcast packet takes the size of 25bytes=200bits. In GCC the coupling strength uses the empirical value $\epsilon = 0.4$. Figure 4 shows the comparison of energy consumption ratio of the three clustering algorithms from sparse to dense network environment. The energy consumption is calculated using the first order energy model and for each node the total energy is set as $2J$. All the results data appeared in this figure are averaged over 500 network topologies and initial phases realization.

We use the most simple version of LEACH, in which the cluster head nodes are chosen based only on the rotation probability. Neither the residual energy nor other requirements are taken into consideration. In reality it will cost more than that but from the energy perspective, LEACH is still superior to HEED and LEACH because of its simplicity. For GCC, energy cost increase along with the node density and the different delay values almost have no influence. While for HEED, energy cost linear increase with the node density, and bigger transmission radius r cost more energy. This is because HEED chooses best suitable node (highest residual energy, min-cost) as head node from its neighbors. So when the density increase or neighborhood enlarges, it has to communicate and check more neighboring nodes. When transmission radius $r = 30$ and $\tau = 0.05$, HEED and GCC both produce around 10 clusters. Every node in GCC cost more energy than in HEED. At $r = 40$ and $\tau = 0.1$, they both produce around 5 clusters. In sparse network GCC still cost more than HEED but as the node density increase, GCC's energy consumption ratio gets close to HEED and become lower than HEED when $n > 300$. Therefore, the trend is when the network gets more dense, the GCC's energy performance becomes better.

From the comparison we can see GCC's energy cost is acceptable for once cluster construction(below 0.1%). But if we consider GCC's much longer reconstruct period and the extra sync/desync benefit, the total algorithm energy consumption is superior to the normal clustering algorithms.

VI. CONCLUSION AND FUTURE WORK

The proposed GCC method builds connection between application level requirements and clustering constructions by taking natural characteristics of the computing nodes as general clustering criteria. At the same time of clustering, the method also achieves synchronization inside a cluster and desynchronization between clusters. It is not only highly efficient but also provides a foundation to our future work, in which the data gathering framework can be constructed and the data aggregation and desynchronized transmission schedule techniques will be used.

ACKNOWLEDGMENT

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National Industry Promotion Agency) (NIPA-2010-(C1090-1021-0003)).

REFERENCES

[1] R. Mirollo and S. Strogatz, "Synchronization of pulse-coupled biological oscillators," *SIAM Journal on Applied Mathematics*, vol. 50, no. 6, pp. 1645–1662, Dec 1990.

[2] Y. Niu, B. J. d'Auriol, X. L. Wu, J. Wang, J. S. Cho, and S. Y. Lee, "Selective pulse coupling synchronicity for sensor network," in *Second International Conference on Sensor Technologies and Applications 2008, (SENSORCOMM '08)*, Cap Esterel, France, Aug 2008, pp. 123–128.

[3] G. Werner Allen, G. Tewari, A. Patel, M. Welsh, and R. Nagpal, "Firefly inspired sensor network synchronicity with realistic radio effects," in *Proc. 3rd ACM Conference on Embedded Networked Sensor Systems (SenSys'05)*. San Diego, California, USA.: ACM Press, Nov 2005, pp. 142–153.

[4] A. Tyrrell, G. Auer, and C. Bettstetter, "Fireflies as role models for synchronization in ad hoc networks," in *Proc. of 1st Bio-Inspired Models of Network, Information, and Computing Systems 2006, (BIONETICS)*, Madonna di Campiglio, Dec 2006, pp. 1–7.

[5] U. Ernst, K. Pawelzik, and T. Geisel, "Delay-induced multistable synchronization of biological oscillators," *Phys. Rev. E*, vol. 57, no. 2, pp. 2150–2162, Feb 1998.