Efficient Multi-level Clustering for Very Large Wireless Sensor Networks with Gateways Support and Meta-heuristic Integration

Basilis Mamalis
University of West Attica Ag. Spyridonos, 12243, Athens, Greece

Sotiris Mamalis
National Bank of Greece Aiolou 86, 105 59 Athens, Greece

Marios Perlitis
Democritus University of Thrace, Campus, 69100, Komotini, Greece

ABSTRACT
Node clustering is an efficient technique for saving energy of wireless sensor networks (WSNs). In this paper we present a novel hybrid multi-level clustering scheme that combines a traditional, appropriately modified, gradient-based clustering protocol with an evolutionary optimization method that is mainly based on the Gravitational Search Algorithm (GSA). The proposed scheme aims at improved performance over large in size networks, where classical schemes in most cases lead to non-efficient solutions. It first creates suitably balanced multi-hop clusters, in which the sensors energy gets larger as coming closer to the cluster head (CH). This scheme is then extended to generate a hierarchy of cluster-heads with the same characteristics; note that the energy savings increase with the number of levels in the hierarchy. In the last level of the proposed scheme a suitable protocol based on the GSA runs to associate sets of top-level cluster-heads to specific gateway nodes for the eventual relaying of data to the base station (BS). The fitness function was appropriately chosen considering both the distance from the cluster heads to the gateway nodes and the remaining energy of the gateway nodes, and it was further optimized in order to gain more accurate results for large instances. Extended experimental measurements demonstrate the efficiency and scalability of the presented approach over very large WSNs, as well as its superiority over other known clustering approaches of the literature, with the same objectives.

General Terms
Networks, Sensors, Algorithms

Keywords
Wireless Sensor Networks; Gravitational Search Algorithm; Node Clustering; Network Lifetime; Energy Balancing

1. INTRODUCTION
Grouping sensors in clusters is an effective method for saving energy in large-scale WSNs [1]. Considering such a WSN, there are several sets of sensors called clusters and each one of them has a leader called 'cluster head'. The sensor nodes (after sensing the field) send the sensed data to the CH, and then the CH (after collecting the data) forwards them to the BS. A study in WSN clustering is given in [1]. Furthermore, a lot of scientists have adopted the use of a specific kind of nodes called 'gateway-nodes' which operate similarly to the normal sensor nodes, however they are usually equipped with more energy and communication capabilities and they cost more [2]. The gateway nodes finally behave as CH in the WSN, and forward the gathered data to the BS. Actually the

scope of this idea is to create a stronger group of CH than in typical networks with cluster organization. On the other hand, their operation is still based on batteries, so they need to preserve their energy adequately while the network operates.

Let's suppose we have m gateway nodes and n sensors in the network, and there are k nodes in the range of the gateway nodes. The complexity for associating the n sensors to the m gateway nodes is naturally exponential. So, typical techniques can't normally lead to effective solutions that scale well. A lot of evolutionary methods could be applied to give a good approximate solution. The GSA [3] has been recently reported as a quite valuable such technique, driven by nature, which can manage effectively problems in hard computational form. Also, the particle swarm optimization technique and the ant colony method have been thoroughly studied [1].

A large number of research works can be found in the bibliography [1,4-17,22-32] following clustering schemes in WSNs (either centralized or distributed). The LEACH protocol [8] is the most known and one of the most effective. In this protocol the cluster-heads are chosen on the basis of a specified probability. The disadvantage of this protocol is that a sensor node of less energy could be chosen as a cluster head, thus limiting the network operation. PEGASIS [9] was proposed as an efficient alternative, but it can't lead to good scalability because of several constraints. A number of other schemes adopting clustering have further been met, trying to achieve sufficiently balanced depleting of energy and thus improving the network operation. Many of them [4,22-23,31-32] aim at solving the 'hot-spot' problem and achieving ideal distribution in the depletion of energy for all the nodes, whereas others [28-30] try to gain efficiency and scalable secure behavior when dealing with large-scale networks and deployment areas. A comprehensive analysis including the integration of clustering in modern data gathering schemes may also be found in [33].

Focusing on the schemes based on the use of 'gateway nodes', in [13] the authors propose a scheme of satisfactory balance, named LBC. In LBC the clusters are formed quite efficiently, but the algorithm doesn't take in account neither the distance between the sensor nodes and the gateway nodes nor the remaining energy of the gateway nodes. In [14] the authors propose an approach named GLBCA applying BFS, but when large-scale WSNs are considered non-satisfactory execution times are observed. In [15] the authors present an approach based on GA-clustering in order to elect a set of cluster heads among the regular sensors. In [16] the authors propose a scheme based on PSO by taking in account the energy and the

distance within the clusters as main criteria, but on the other hand they don't consider the energy and the distance of the gateway nodes.

In [17], a clustering scheme of high efficiency is presented built over the GSA approach, in which the fitness function is properly selected considering not only the distance among the sensor nodes and the gateways/BS, but also the remaining energy of the gateway nodes. The eventually proposed scheme, named GSA-EEC, is evaluated based on several metrics, and it is proved to be very effective and performing better than most of the related approaches [13-16]. On the other hand, when large-scale networks are considered, these methods in most cases can't offer satisfactory results because of various problems met in the case of very high number of dimensions in the search space. More concretely, the GSA method in such cases is shown to have limited stability as well as low levels of accuracy due to the increased possibility to be trapped into local optimum solutions; additionally non-efficient execution times are observed. The above shortcomings restrict the worth of using the GSA technique in very large networks. The GSA approach has also been used in WSNs for nodes localization and relay/sink nodes placement, as shown in [18-21].

Trying to overcome the disadvantages referred in the previous paragraph with regard to the use of the GSA-oriented scheme [17] on large networks, we've designed a novel hierarchical clustering protocol that consists of multiple levels, and combines adequately the above GSA-oriented scheme with a classical gradient clustering scheme, similar to [22-23] and suitably modified in order to generate less number of clusters and fit better in the energy balancing needs of our proposed total scheme. The presented approach (which extends and improves adequately the work of [24]) first constructs multi-hop clusters of evenly distributed energy reserves, in which the sensor nodes energy gets larger as coming closer to the CH. Thus, since the sensors near to the CH are more exhausted because of data forwarding, keeping the residual energy of those sensors high guarantees the prolonged and seamless operation of the WSN. The above (first level) clustering scheme is then extended to generate a hierarchy of CHs (in an number of intermediate levels) with the same characteristics. As a result the energy savings increase with the number of levels in the hierarchy. Finally, a suitable protocol based on the GSA runs to associate sets of the top-level CHs of the hierarchy to specific gateway nodes for the eventual forwarding of data to the base station. Extended simulated experiments (with Castalia simulator [25]) are presented to show the high efficiency and scalability of the proposed combined multi-level clustering scheme over very large WSNs, as well as its superiority over other clustering approaches of the literature [26-27,31,35].

The primary goal of our work is to build a clustering protocol able to achieve highly efficient behavior in very large-scale WSNs (i.e. not only for hundreds or one-two thousands of sensor nodes but also for 5000-10000 and even more sensors; as it will be the case in future IoT/IoE applications). The existing clustering protocols unfortunately suffer from several shortcomings that lead to performance degradation in such very large scale WSNs, either relevant to the hot spot problem or to the fact that in order to face the hot spot problem they tend to generate quite large number of CHs and/or multi-hop routing overhead when we have large numbers of sensors (making their further use problematic, i.e. for data gathering), and they generally don't lead to quite scalable solutions. A good solution can be given by combining clustering protocols

with specific features into suitable multi-level hierarchical schemes trying to gain efficiency and extent the scalability of the total scheme. The main contribution of our work is in the above direction, whereas the proposed scheme also achieves significant improvements against [35], which is a previous work of ours, using a restricted two-level only clustering scheme, with similar objectives.

The remaining text is organized as follows. In section 2 some background is firstly given with respect to the GSA technique and multi-level clustering setup. Then, in section 3 our complete multi-level clustering scheme is described, along with proper explanations. In section 4 our extended simulated experiments are presented and the corresponding results are thoroughly discussed, whereas section 5 concludes the paper.

2. MATERIALS AND METHODS
2.1 The gravitational search algorithm
The description of the GSA technique in details is given in [3]. Let's assume we have a group of agents, N_A. Every agent is expected to give a part of the solution. The location of agent A_i, $1 \leq i \leq N_A$ in dimension d is x_i^d whereas its velocity is v_i^d, $1 \leq d \leq D$. Every agent has the same dimension. Every agent is evaluated to verify the suitability of the result based on a specific fitness function. Let the ith agent's location be represented as $X_i = (x_i^1, x_i^2, ..., x_i^D)$. The following expression gives the force applied on the ith agent by the jth agent.

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t)} (x_j^d(t) - x_i^d(t))$$

M_{pi} represents the passive mass of the ith agent, whereas M_{aj} stands as the corresponding active mass of the jth agent, and α is a constant. $G(t)$ equals to $G_0(t_0/t_{max})^\phi$, in which G_0 represents also a constant. $R_{ij}(t)$ means the Euclidean distance from agent i to agent j. The overall force applied by the group of agents over the ith agent in dimension d is as follows.

$$F_i^d(t) = \sum_{j=1, j \neq i}^{N_A} rand_j \times F_{ij}^d(t)$$

The inertial mass as well as the gravitational mass are estimated by evaluating the fitness function. The heavier the mass of an agent the more efficient the agent is.

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} + \varepsilon$$

In the above, ε is a limited constant, $fit_i(t)$ represents the fitness of agent i, whereas *worst (t) / best (t)* are as next.

$$best(t) = \min_{j \in \{1..N_A\}} fit_j(t)$$

$$worst(t) = \max_{j \in \{1..N_A\}} fit_j(t)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N_A} m_j(t)}$$

Further, let's assume M_{ai}, M_{pi}, M_{ii} and M_i are all equal to each other. Based on the 2nd law of Newton, we conclude to the next formula.

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)}$$

In the above, the inertial mass of agent i is given by M_{ii}, whereas the acceleration of agent i is given by and $a_i^d(t)$.

2.2 Multi-level clustering setup

As previously stated, the proposed approach has been developed on the basis of a hierarchical clustering protocol of multiple levels, named GC-GSA. First, we assume that the sensors and the gateway nodes spread at random over a large deployment area. The deployed sensors / gateways are also supposed to be static. Then, a two-phase network operation begins. In the first phase the necessary bootstrap and cluster formation procedures (for all the multiple levels) are completed. Initially, proper identities are assigned to all the sensors by the base station. Then, both the sensors and the gateway-nodes send their identities and other info to their neighbors using mac-level protocol. The gateway-nodes are finally informed with the identities of their neighbor sensors. Next, every gateway-node informs the base station with the collected info to complete the setup. The complete multi-level clustering routine is then run and all the sensors nodes get the necessary info with regard to their local CH. In the second phase the network starts to operate steadily; the sensed data are gathered by the top-level CHs and then by the gateway-nodes, and they finally directed to the base station.

3. CLUSTER FORMATION

3.1 First Level Energy-Balanced Clustering

With regard to the initial clustering, a suitable protocol based on multiple hops communication is adopted, aiming at building a robust hierarchy with controlled delay and sufficient coverage with respect to the secondary clustering (e.g. the size of the clusters and the positions of the selected cluster-heads). Further, the algorithm of [4] is primarily followed as the firstly applied clustering routine, whereas a proper adaptation has been incorporated (mainly in step 3) aiming at inheriting the main advantages of the approach described in [22,23]. The remaining energy of the sensors is chosen as the basic factor while building the clusters, so a sufficiently balanced cluster hierarchy is built, the energy-hole problem (near to the cluster-heads) is effectively handled, and eventually the operation of the network is prolonged. Specifically, our initial (first-level) cluster formation algorithm consists of the following basic steps:

1. Initially, all the sensors in the network broadcast messages (including their residual energy and their ID) in a certain power, which ensures that nodes within a radius r will receive the message. Then, each node waits to receive such messages from all its 'neighboring' nodes.

2. For every such received message, each node compares the residual energy in the message with its own energy, and then, if the energy in the message is larger it marks the node which sent the message as its new parent.

3. If the node which received the message (let's name it v) has already a parent, and the node which sent the message has larger residual energy than its own (i.e. it's also a candidate parent), it computes the ratio '$E_{diff}(v)$/Dist(v)'[1] for both its parent and the new candidate parent, and it chooses as its next parent the one who maximizes the above fraction.

4. When a node has received all its neighboring nodes messages and has made the necessary decisions, it sends a final 'join' message to its parent node, and marks itself as a 'member' node for the rest of the protocol. If the node has no parent node, then it marks itself as a CH and broadcasts a relevant message.

Thus, at the end of execution each node has as 'parent' a node that has quite larger residual energy than its own, in a sufficiently close distance too. Experimental results in [4,22-23] show that the above kind of clustering algorithms effectively save the energy costs, lead to balanced energy consumption and prolong the lifetime of the network. Furthermore, the main goal of such algorithms is the creation of suitable (energy-balanced) clusters with not only high-energy CHs, but also having energy-rich neighborhoods. In that way they effectively avoid energy holes around the CHs and naturally they become quite suitable for initial (first-level) clustering (as well as for energy-efficient data gathering using a MS, as shown and discussed in [22-23,34] in more details).

On the basis of the above considerations, we have followed the use of the proposed modified algorithm as the first-level clustering protocol (as opposed to the use of the relevant watershed-based algorithm of [22-23] or the use of the much simpler but weaker protocol of [4]), mainly because it has the potential to perform and scale better for large and very large / huge numbers of sensor nodes and relevant deployment areas, in the context of this problem.

3.2 Intermediate Levels Clustering

The first level clustering algorithm described in the previous paragraph is then extended to allow more than one levels of clustering. Assume that there are h levels in the clustering hierarchy with *level 0* being the lowest level and *level h-1* being the highest. In this clustered environment, the sensors communicate the gathered data to *level-0* CHs. The *level-0* CHs communicate the gathered data to *level-1* CHs and so on. Finally, the *top level (h-1)* CHs communicate the gathered data to the 'gateways' they have been assigned to, and the gateway-nodes forward the data to the base station.

More concretely, the proposed hierarchical cluster formation algorithm works in a bottom-up fashion. The algorithm first elects the *level-0* CHs, then *level-1* CHs, and so on. The *level-0* CHs are chosen as described in par. 3.1, i.e. by applying the first-level energy balanced clustering method over all the sensor nodes. In the next phase, the same clustering algorithm is applied over the *level-0* CHs to elect the *level-1* CHs, with the effective transmission range (i.e. the transmission range used for control messages) properly adjusted. The latter is necessary to assure that sufficient number of *level-0* CHs will be in the transmission range of each other, and the clustering procedure will act in the proper way. *Level-2* CHs are elected in the same way, and so on, until the number of the *top-level* CHs become small enough, and the last-level GSA oriented gateways assignment procedure (see par. 3.3) may be applied effectively. In Fig. 1 a sample view of the final cluster hierarchy is given. It has been generated for *h=3* cluster levels and it shows (as a typical example and without loss of generality) the construction of one such *top-level* CH which is then assigned to one of the available gateway-nodes of the network.

---

[1] Where '$E_{diff}(v)$' is how much more energy has the candidate parent node than node v, and 'Dist(v)' is the Euclidean distance between node v and the candidate parent node.
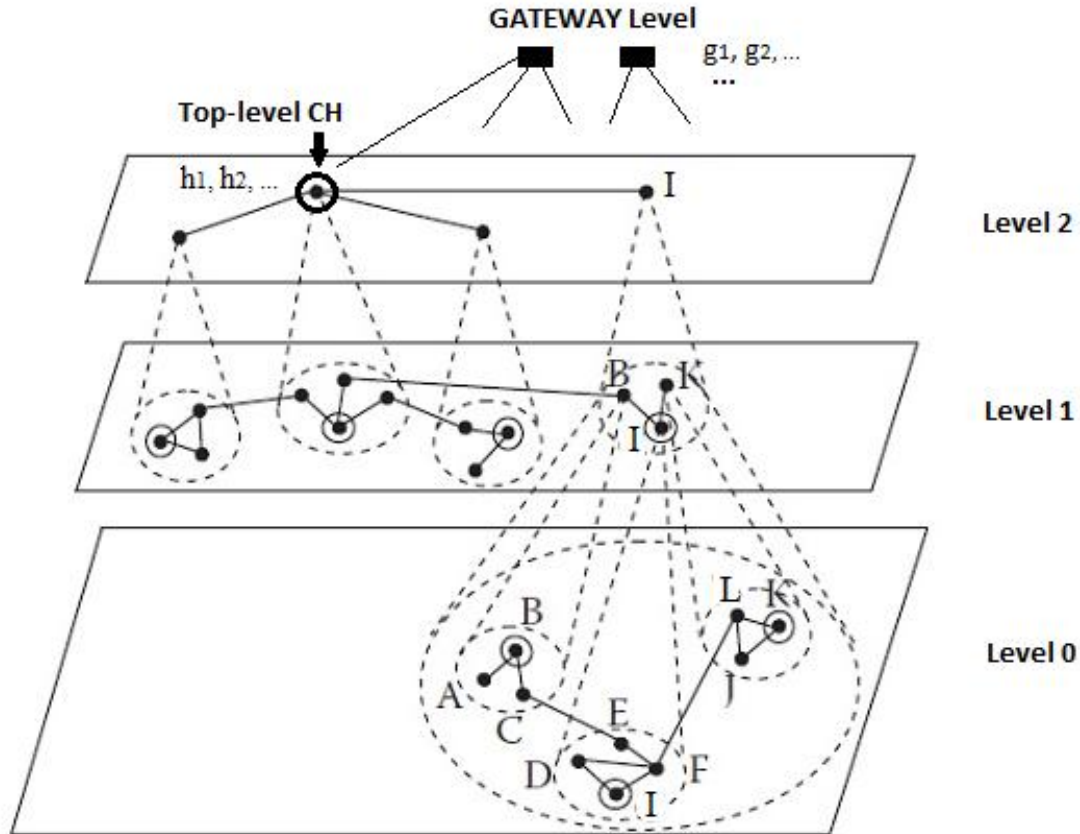
Fig 1. An example of the multi-level cluster hierarchy

3.3 Last Level GSA-based Clustering

Considering the last level of building the complete cluster hierarchy (including the gateway nodes), a GSA-based protocol has been adopted. The design of our protocol has been mainly based on the approach of [17] which has been suitably modified and applied over the total set of the 'top-level' CHs elected in the previous phases. The fact that the set of the elected top-level CHs is of relatively small size allows the effective use of faster centralized evolutionary algorithms with high accuracy. In this context, the proposed GSA oriented scheme runs to associate sets of top-level CHs to specific gateway nodes for the eventual relaying of data to the base station. Since prolonging the network lifetime remains the main priority, the fitness function was further optimized and modified in order to keep the energy balanced over the gateway-nodes too, and gain more accurate results for large and very large / huge instances.

More concretely, the initialization of the basic group of agents (N_A) first takes place (note that each agent potentially stands as a part of the solution). Specifically, in the context of this stage of clustering the agents represent the associations of cluster-heads to corresponding gateway-nodes. If we assume that A_i stands as the agent i, every item $x_i^d(t)$ associates the relevant cluster-head to some gateway and $1 \leq i \leq N_A$, whereas also $1 \leq d \leq D$ (note that D equals to c). So, each agent may be denoted in the following form [17,33], whereas our modified GSA protocol should then run as follows.

$$A_i = [x_i^1(t), x_i^2(t), x_i^3(t), ..., x_i^D(t)]$$

The Gateway-level GSA-based Clustering Algorithm

Input:

– Group of (top-level) CHs: $H = \{h_1, h_2, h_3, ... , h_c\}$

– Set of gateway nodes: $G = \{g_1, g_2, g_3, ... , g_m\}$

– Initial group of agents with size equal to N_A

– Agent's dimensions = # of (top-level) CHs = c

Output:

The optimal (top-level) CHs associations to gateway-nodes

Description of the algorithm:

Agent A_i is initialized, $\forall i, 1 <= i <= N_A$

The mapping function is defined (for every h_d to a g_k)

do /* *initially assume that t=0* */

  for *i=1 to N_A*

   Fitness (A_i) is computed

   The *best/worst* fitness values are updated of all agents

   $M_i(t), a_i^d(t)$ are computed of each agent of the system

   The velocity and the position of A_i are updated

  endfor

while the criteria for termination are not satisfied

3.4 Choosing the Fitness Function

The definition of fitness function (*f*) has to be suitably specified considering not only (i) the residual energy of the gateway-nodes, but also (ii) the distance between the top-level CH and the gateway-node as well as between the latter and the base station. The gateway-nodes of high remaining energy reserves should be elected. Thus, the energy consumption is suitably balanced and the lifetime of the network is prolonged. Moreover they should be the ones having the less distance too. These requirements could be further described as follows (E_{gj} stands as the remaining energy of g_j, and l_{ij} equals to the number of gateway-nodes in the neighborhood of h_i).

Fitness Function:

Object. 1: *Maximize f1* = $\sum_{i=1}^{c} E_{gj}$

Object. 2: *Minimize f2* = $\sum_{i=1}^{c} (d(h_i, g_j) + d(g_j, BS))$

Total Object.: Minimize $f = \dfrac{\beta \times f2 + t_1}{\alpha \times f1 + t_2}$

Through the above function we propose an optimized approach to the one of [17], aiming at the maximum possible balance on the actual behavior of the two main factors. Note that α, β and t_1, t_2 are residual-energy/distance dependent and independent constants respectively, whereas the final value of the function is restricted adequately between 0 and 1 to normalize and optimize the result. The new proposed fitness function not only balances the weight of the two main factors (energy and distance) in the final computation, but also allows the designer to normalize (through α, β) conveniently the unpredictable (non-canonical) gaps caused by the potentially different measure units. It gives also the flexibility to take in account (through t_1, t_2) during the optimization process other significant parameters too (like data rates, transmission range, initial energy and energy consumption rates etc.).

3.5 Discussion and Extensions

As it comes out of the literature, almost all the existing clustering protocols suffer from several shortcomings when applied over very large scale WSNs (either relevant to the *hot spot* problem or to undesired routing overheads, and generally lead to not sufficiently scalable solutions. As a consequence a proper solution could be given by combining appropriately clustering protocols with specific features into corresponding *multi-level* hierarchical schemes trying to gain efficiency and extent the scalability of the total scheme. In the above context, with the use of the GSA oriented scheme as the upper level cluster formation routine, we get over the related disadvantages, due to the fact that it operates on a quite restricted number of nodes/cluster-heads. Thus, we may gain from its efficient behavior over such WSNs (of restricted size), and finally get outstanding performance gains if we use it in large / huge WSNs together with a lower level cluster formation scheme of similar efficiency (considering the balance of energy consumption), like the protocol used for the first and intermediate levels clustering in our case. Specifically, as it is also referred in section 4, the GSA oriented scheme behaves with great efficiency when applied over a set of e.g. up to 100-500 sensors. As a consequence, our combined approach may behave similarly over e.g. up to 100-500 cluster-heads, thus having the potential to lead to excessive total performance for large / huge networks of 5000-10000, 20000 and even more sensors.

4. SIMULATION RESULTS

In this section, our extended simulation experiments are presented, in which the proposed approach is shown to achieve high efficiency and scalability. Also our approach is compared to the original GSA oriented scheme given in [17] without any modifications. Specific sets of experiments have also been executed to show the suitability of the newly designed fitness function, as well as its superiority over other multi-level / large-scale clustering protocols of the literature [26-27,31,35]. The experiments have been completed with the use of Castalia (a WSN simulator based on OMNeT++ [25]), and they focus on measuring the corresponding in each case protocol's efficiency (measuring the consumption of energy and lifetime of the WSN), over very large networks and deployment areas. In order to have a right comparison, the radio model used is the same as in [8,17].

4.1 Comparing to the native GSA scheme

Trying to compare the proposed combined scheme to the original GSA oriented scheme of [17], specific experiments have been run, considering different numbers of sensor nodes (*n* equal to 500, 1500, 2500), deployed at random in a grid area with its side taking values between 200m and 500m (i.e. deployment areas of 200x200m^2, 360x360m^2, 480x480m^2). The two first parameter values (500, 200x200m^2) represent the main setup size chosen in the experimental evaluation of [17]. Based on the remaining two setup parameter values an approximate scale of 3x and 5x is simulated with respect to the experiments of [17]. With regard to the network setup factors considered in our simulations we have used the same values as given in the experiments of [17], mainly for comparison reasons. We've also defined 0, 1 or 2 intermediate levels in our cluster hierarchy depending to the value of *n*. The corresponding results are given in Fig. 2-5. By observing Fig. 2-4 one can easily conclude that considering the consumption of energy for each node, our proposed approach is definitely superior to the native GSA-EEC approach when n is equal to 1500 and 2500. On the other hand, it achieves slightly worse performance when n is equal to or less than 500 nodes. Specifically, for n equal to 1500, GC-GSA is shown to have a decrement of 16,5% with respect to the energy consumed on average, and this decrement becomes equal to 28% for n equal to 2500 sensor nodes. Instead, GSA-EEC has an almost ideal behaviour when n is equal to 500 sensor nodes, and its behaviour becomes gradually worse and worse as the WSN increases in size. Moreover, Fig. 5 presents the lifetime of the network with respect to the various test scenarios addressed above. One may easily observe that the GC-GSA approach leads to significant increments of the WSN lifetime when considering large networks (up to 30% - and even more - when n is equal to 2500 nodes).
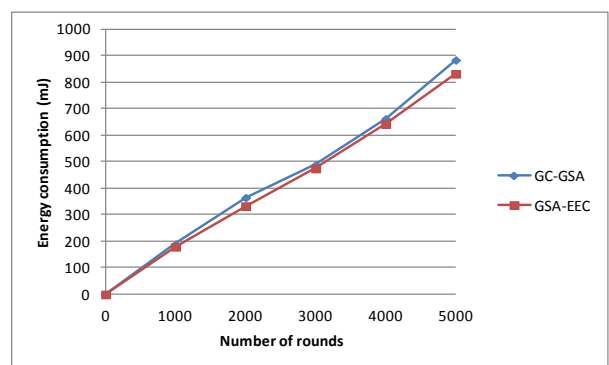

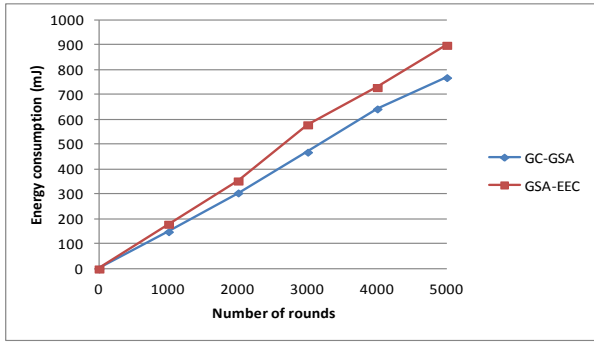**Fig 2. Average energy consumption for n=500**
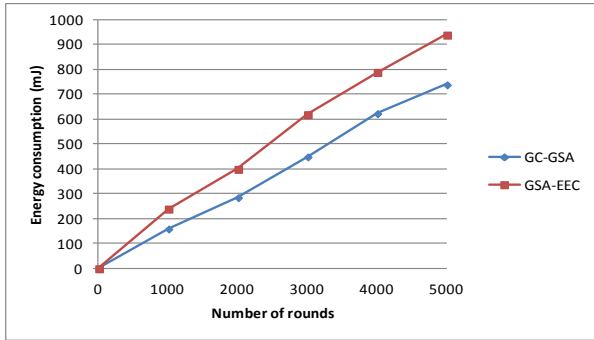
Fig 3. Average energy consumption for n=1500

Fig 4. Average energy consumption for n=2500
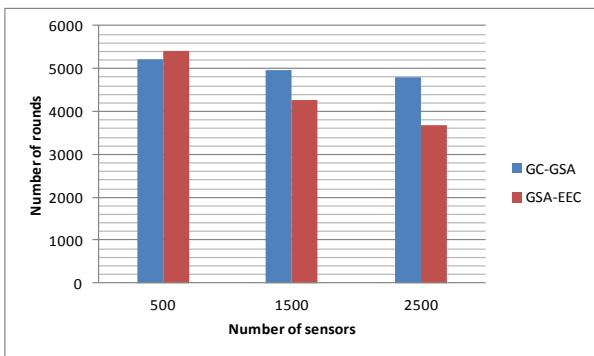

Fig 5. Network lifetime for varying 'n'

4.2 Evaluating the scalability of GC-GSA

Moreover, in order to further examine the high efficiency and scalability of our combined protocol (GC-GSA algorithm) with use of the optimized fitness function of section 3.4, we've run additional experiments for very large WSN deployments - up to n = 10000 sensors. Specifically, we've examined the behavior of the GC-GSA protocol (in terms of energy consumption and network lifetime) for 4000, 6000, 8000 and 10000 sensors, over corresponding progressively growing deployment areas – from 800x800m^2 to 2000x2000m^2 (i.e. 800x800m^2,1200x1200 m^2,1600x1600m^2 and 2000x2000m^2 terrains). We've also defined the use of 2 or 3 intermediate levels in the proposed cluster hierarchy, depending to the value of n. Moreover, we've used as the base for comparison the network lifetime achieved for 1500/2500 sensor nodes, which is approximately equal (see also Fig. 5 as well as the relevant discussion) to 5000 rounds.

Furthermore, we've also chosen progressively growing deployment areas in order to test our approach in more realistic/practical cases; note here that if the size of the deployment area had been kept the same (i.e 480x480m^2 as it was for 2500 sensors) our combined approach would normally

lead to almost equal measurements, since our base-level clustering protocol is not influenced significantly (by its nature - see also [22-23] with respect to the quite similar watershed-based clustering protocol) by the increased density of sensors within the deployment area.

As it is shown in Fig. 7 there is a progressive decrease in the network lifetime when the number of the deployed sensors increase, which is more clear/significant for 10000 sensor nodes (over a 2000x2000 terrain). More concretely, the corresponding decrease for 4000, 6000, 8000 and 10000 sensors is approximately equal to 1.5%, 5,2%, 8.5% and 11.5% respectively. This decrease may not be considered insignificant, (at least for 8000 and 10000 sensors), however it is quite expected due to the large extent of the corresponding deployment areas. Due to the progressively growing deployment area the number of CHs and their sizes increase with a much less structured way, thus making much more difficult to keep the desired balance in energy consumption.

The observed decrease in the network lifetime is mainly caused by the relevant increase in the energy consumed by the sensor nodes (in average), as it is shown in more details in Fig. 6. More concretely, one can easily observe (staring specifically at the curves for 8000 and 10000 sensors) that the average energy consumption increases slightly till the execution of approximately 4000 rounds (due to the reasons referred above), and progressively more sharply afterwards, as the result of the energy exhaustion of some sensors and the end of network lifetime in every case. Overall, we can say that the proposed GC-GSA algorithm scales quite well even for very large number of sensors and deployment areas, thus making itself a promising choice for such extent realistic applications. Moreover, as it was also discussed in subsection 3.5, it can be easily combined and integrated with one or more (mobile or not) sinks/collectors, thus forming a robust and flexible data gathering solution for very large-scale WSN based application environments.
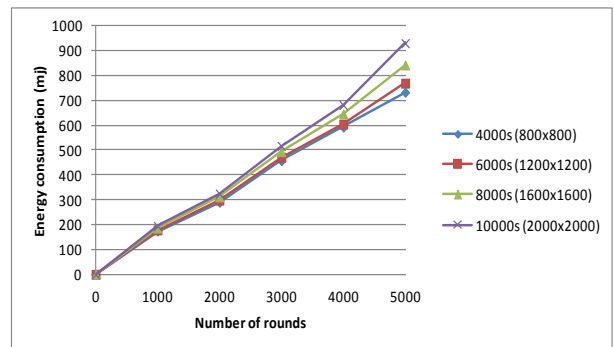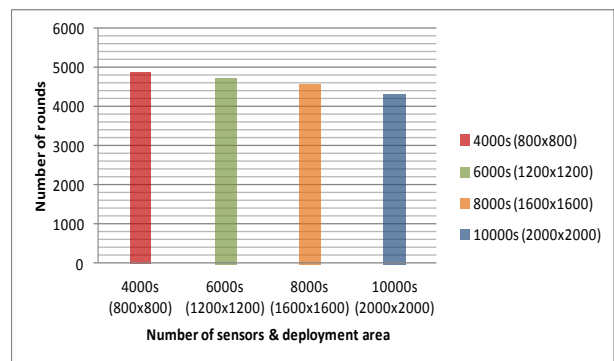

Fig 6. Average energy consumption for very large WSNs

Fig 7. Network lifetime for very large WSNs

4.3 Comparing to other approaches

Furthermore, in order to have a more clear sense of the performance of our multi-level clustering protocol we've performed additional experiments comparing the proposed approach to other clustering approaches of the literature, such as (a) TL-LEACH [26] and EEHC [27], which are two of the most known multi-level clustering protocols, (b) the protocol presented in [31] which is one two of the most valuable recent clustering approaches for large-scale WSNs, and (c) the work of [35] which is a previous work of ours, using a similar (but restricted in one base level only) node clustering scheme. Specifically, we've run experiments for n=2500 sensors (and 2 intermediate levels in the cluster hierarchy) over a 480x480m2 terrain (which is a representative case of a very large WSN instance), and we've compared the performance of all the protocols over that instance. The corresponding results with regard to the energy consumption of all protocols in average, are given in Fig. 8. In Fig. 9, the lifetime of the network for all the algorithms is shown (with additional measurements taken for n=1500 sensors and 1 intermediate level over a 360x360m^2 for comparison reasons).

In [26] the authors propose a hierarchical protocol that consists of two levels of clustering (TL-LEACH), aiming at managing the consumed energy in a more efficient way. The protocol rotates both the first-level and second-level cluster-heads at random. In this way a two-level hierarchy is built, where it is possible, thus leading to a more efficient distribution of the energy reserves between the sensor nodes. Moreover, it is shown to perform much better than LEACH in terms of energy consumption and network lifetime. In [27], a random cluster-based hierarchy is proposed to organize the sensor nodes in a distributed manner. Their basic algorithm is then extended (EEHC) towards the construction of additional levels of cluster-heads, and finally a significant increase is observed with respect to the consumed energy. In [31], the efficient solution of the 'hot spot' problem is the main objective of the authors. Their effort is based on the suitable rotation of the role of cluster-head between al the sensors, as well as adjusting appropriately the size of the formed clusters. The proposed protocol (UCF) first selects as probable CHs the sensors that have more residual energy in the local area. Next, a fuzzy-logic technique is employed for the adjustment of the radius of the cluster. Simulation results show that the above protocol achieves significant improvements in the basic performance metrics. Finally, in [35] a relevant work of ours is presented, introducing a similar cluster formation scheme that preserves energy balancing at the base level only, and leads to very good results for large-scale WSNs.

As it is shown in Fig. 8 the energy consumption of the GC-GSA algorithm is less in average than all the competing protocols. The network lifetime (Fig. 9) is also steadily kept in high levels (around 5000 rounds), and its superiority over the other protocols is clear for both 1500 and 2500 sensors. The performance of TL-LEACH [26] is quite satisfactory, especially for 1500 sensors. However, although it also uses (among else) localized coordination to enable scalability, it employs a probability model for CH selection and so its energy efficiency can't be maximized. As a result it cannot preserve competitive efficiency comparing to GC-GSA for large and very large number of sensors. The EEHC protocol has similar behaviour with TL-LEACH since it's a randomized approach too; Moreover, it performs slightly better due to its modular nature and the fact that it pays more attention in energy efficiency than TL-LEACH (stochastic geometry techniques). However for the same reasons it's also

not highly competitive comparing to GC-GSA for large and very number of sensors. The best performance among the other protocols is given by the UCF protocol [31], as well as by our previous approach presented in [35]. They both lead to similar (slightly worse) network lifetime and average energy consumption with the multi-level GC-GSA. The fact that no other factors than the residual energy are taken in account for CH selection in [31], as well as the uncertainty nature of the fuzzy logic procedure makes UCF not scaling the same well for very large numbers of sensors (1500 and 2500 sensors) as for smaller in size networks. On the other hand, the fact that our previous approach presented in [35] restricts the cluster hierarchy to two levels only (attempting energy balancing over a huge single/base level area), leads practically to similar behaviour (slightly better) as UCF. As a result it cannot scale the same well for very large numbers of sensors, since its energy balancing structure becomes less tight as the network size and the deployment area increase.
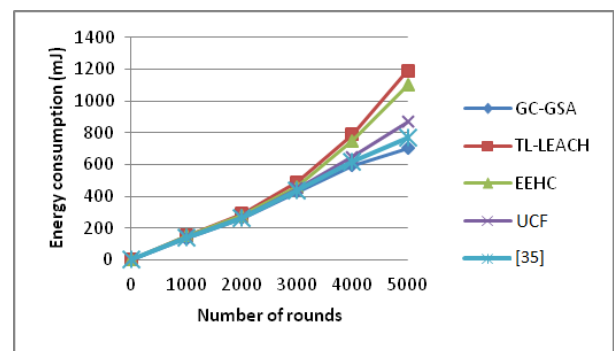

Fig 8. Average energy consumption for GC-GSA, TL-LEACH, EEHC, UCF and [35] (for 2500 sensors)
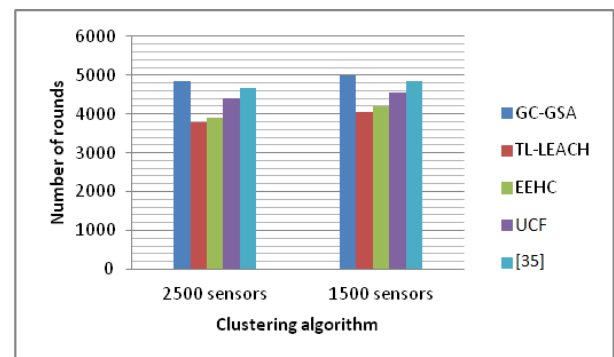

Fig 9. Network lifetime for GC-GSA, TL-LEACH, EEHC, UCF and [35] (for 1500 and 2500 sensors)

5. CONCLUSION

Throughout this paper the worth of using a novel hybrid clustering scheme that completes in multiple hierarchical phases is outlined. The proposed approach first constructs multi-hop clusters of evenly distributed energy reserves, in which the sensor nodes energy gets larger as coming closer to the cluster head. Thus, since the sensors near to the cluster head are more exhausted because of data forwarding, keeping the residual energy of those sensors high guarantees the prolonged and seamless operation of the WSN. The above scheme is then extended to generate a hierarchy of cluster-heads with the same characteristics. Finally, a suitable protocol based on the GSA runs to associate sets of cluster heads to specific gateway nodes for the eventual relaying of data to the base station. A fitness function was appropriately chosen and optimized considering both the distance from the

cluster heads to the gateway nodes and the remaining energy of the gateway nodes. Extended simulated experiments have been completed to show the high efficiency and scalability of the proposed scheme over very large WSNs, as well as its superiority over other known approaches of the literature, including a relevant work of ours of very promising behavior with a two-level hierarchy.

6. REFERENCES

[1] B. Mamalis, D. Gavalas, C. Konstantopoulos, and G. Pantziou. 2009. Clustering in wireless sensor networks. Book chapter. In RFID and sensor networks: Architectures, protocols, security & integrations, New York: CRC Press, Chap. 12, pp. 324–353.

[2] P.C.S. Rao, H. Banka, and P.K. Jana. 2016. PSO-based multiple-sink placement algorithm for protracting the lifetime of wireless sensor networks. In AISC, Springer, Heidelberg, vol. 379, pp. 605–616.

[3] R. Esmat, N. Hossein, and S. Saeid. 2009. GSA: A Gravitational Search Algorithm. Information Sciences, 179(13), 2232–2248.

[4] X. Bao, L. Liu, S. Zhang and F. Bao. 2010. An Energy Balanced Multihop Adaptive Clustering protocol for Wireless Sensor Networks. In Proceedings of the 2nd IEEE ICSPS Conference, vol. 3, pp. 47-51.

[5] M. Sabet, and H.R. Naji. 2015. A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks. AEU International Journal of Electronics and Communications, 69(5), 790–799.

[6] S.A. Sert, H. Bagci, and A. Yazici. 2015. MOFCA: multi-objective fuzzy clustering algorithm for wireless sensor networks. Applied Soft Computing 30, 151–165.

[7] D.S. Abbasi, and J. Abouei. 2015. Toward cluster-based weighted compressive data aggregation in wireless sensor networks. Ad Hoc Networks.

[8] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan. 2000. Energy efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Hawaii International Conference on System Sciences, p. 10.

[9] S. Lindsey, and C.S. Raghavendra. 2002. PEGASIS: power efficient gathering in sensor information systems. In Proceedings of the IEEE Aerospace Conference, pp. 1125–1130.

[10] P.C.S. Rao, P.K. Jana, and H. Banka. 2016. A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks. Wireless Networks, Springer (online), 1-16.

[11] P.C.S. Rao, and H. Banka. 2017. Energy efficient clustering algorithms for WSNs: novel chemical reaction optimization approach. Wireless Networks, 23(2), 433–452.

[12] P.C.S. Rao, and H. Banka. 2017. Novel chemical reaction optimization based unequal clustering and routing algorithms for wireless sensor networks. Wireless Networks, 23(3), 759–778.

[13] G. Gupta, and M. Younis. 2003. Load-balanced clustering of wireless sensor networks. In Proceedings of IEEE Intl. Conf. on Communications, ICC 2003, vol. 3, pp. 1848–1852.

[14] C.P. Low, C. Fang, J.M. Ng, and Y.H. Ang. 2008. Efficient Load-Balanced Clustering Algorithms for Wireless Sensor Networks. Computer Communications 31(4), 750–759.

[15] S. Hussain, A.W. Matin, O. Islam. 2007. Genetic Algorithm for Hierarchical Wireless Sensor Networks. Journal of Networks 2(5), 87–97.

[16] N.M.A. Latiff, C.C. Tsemenidis, and B.S. Sheriff. 2007. Energy-aware clustering for wireless sensor networks using particle swarm optimization. In Proceedings of the 18th Annual IEEE ISPIInternational Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–5.

[17] P.C.S. Rao, H. Banka and P.K. Jana. 2015. Energy Efficient Clustering for Wireless Sensor Networks: A Gravitational Search Algorithm. In Proceedings of SEMCCO 2015 Conf.,, Springer, pp. 247-259.

[18] R. Krishnaprabha and A. Gopakumar. 2014. Performance of gravitational search algorithm in wireless sensor network localization, in Intl Conf. on Communication, Signal Processing and Networking (NCCSN), IEEE, pp. 1-6.

[19] Zhao Wei-Guo, Yang Shao-Pu, Li Kui and Wang Li-Ying. 2013. Gravitational Search Algorithm for Node Localization in WSN. Information Technology Journal, 12, 5806-5811.

[20] J.M. Lanza-Gutierrez and J.A. Gomez-Pulido. 2017. A Gravitational Search Algorithm for Solving the Relay Node Placement Problem in WSNs, In International Journal of Communication Systems, 30(2), 1-21.

[21] P.C.S. Rao, H. Banka and P.K. Jana. 2015. A Gravitational Search Algorithm for Energy Efficient Multi-sink Placement in Wireless Sensor Networks. In Proceedings of SEMCCO 2015 Conf., Springer, pp. 222-234.

[22] C. Konstantopoulos, B. Mamalis, G. Pantziou, and V. Thanasias 2012. Watershed-based Clustering for Energy Efficient Data Gathering in WSNs with Mobile Collector. In Proc. of Europar Conf., Springer, LNCS 7484, pp. 754-766.

[23] C. Konstantopoulos, B. Mamalis, G. Pantziou, and V. Thanasias 2015. An Image Processing Inspired Mobile Sink Solution for Energy Efficient Data Gathering in Wireless Sensor Networks. In Wireless Networks 21(1), 227-249.

[24] B. Mamalis, and M. Perlitis, "Energy Balanced Two-level Clustering for Large-Scale Wireless Sensor Networks based on the Gravitational Search Algorithm", In IJACSA Journal, Vol. 10, No. 12, 2019.

[25] Castalia: WSNs and BANs simulator. 2007. National ICT Australia. URL: http:// castalia.npc.nicta.com.au/.

[26] Loscri, V & Morabito, G & Marano, Salvatore. 2005. A Two-levels Hierarchy for Low-Energy Adaptive Clustering Hierarchy (TL-LEACH). In Proceedings of IEEE Vehicular Technology Conference, vol. 3, pp. 1809-1813.

[27] Bandyopadhyay, Seema & Coyle, Edward. 2003. An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. In Proceedings of IEEE

INFOCOM Conference, vol.3, pp. 1713-1723.

[28] Tie Qiu, Aoyang Zhao, Feng Xia, Weisheng Si, Dapeng Oliver Wu, Tie Qiu, Aoyang Zhao, Feng Xia, Weisheng Si, and Dapeng Oliver Wu. 2017. ROSE: Robustness Strategy for Scale-Free Wireless Sensor Networks. IEEE/ACM Transactions on Networking, 25, 5, 2944-2959.

[29] M. Dong, K. Ota and A. Liu. 2016. RMER: Reliable and Energy-Efficient Data Collection for Large-Scale Wireless Sensor Networks. In IEEE Internet of Things Journal, vol. 3, no. 4, pp. 511-519.

[30] Z. Xu, L. Chen, C. Chen and X. Guan. 2016. Joint Clustering and Routing Design for Reliable and Efficient Data Collection in Large-Scale Wireless Sensor Networks. In IEEE Internet of Things Journal, vol. 3, no. 4, pp. 520-532.

[31] P Neamatollahi and M Naghibzadeh. 2018. Distributed unequal clustering algorithm in large-scale wireless sensor networks using fuzzy logic. In the Journal of Supercomputing, 74, 6, pp 2329-2352.

[32] A. Abro, D. Zhongliang, K. A. Memon, K. H. Mohammadani, N. ul Ain, S. Memon, I. Memon and M. A. Panhwar, 2019. Minimizing Energy Expenditures using Genetic Algorithm for Scalability and Longlivety of Multi hop Sensor Networks. In 9th.IEEE ICEIEC Conf., Beijing, China, 2019, pp. 183-187.

[33] C. Konstantopoulos, N. Vathis, G. Pantziou, and D. Gavalas. 2018. Employing mobile elements for delay-constrained data gathering in WSNs. Computer Networks 135: 108-131.

[34] B. Mamalis. 2014. Prolonging Network Lifetime in Wireless Sensor Networks with Path-Constrained Mobile Sink. In IJACSA Journal, Vol.5, No.10, pp.82-91.

[35] B. Mamalis and M. Perlitis. 2019. Energy Balanced Two-level Clustering for Large-Scale Wireless Sensor Networks based on the Gravitational Search Algorithm. In IJACSA Journal, Vol.10, No.12, pp.32-41.