One-Step Clustering Protocol for Periodic Traffic Wireless Sensor Networks

Nassima Merabtine[1,2], Djamel Djenouri[3], Djamel-Eddine Zegour[1], Elseddik Lamini[2,4], Rima Bellal[2,4], Imene Ghaoui[4] and Nabila Dahlal[4]

[1]Computer Science Higher School, ESI, ex. INI
[2]Center for Development of Advanced Technologies (CDTA)
[3]CERIST Research Center
[4]USTHB university
e-mail:n_merabtine@esi.dz

***Abstract*—A new centralized clustering protocol for periodic traffic wireless sensor networks is proposed in this paper. The proposed protocol uses one-step off-line cluster computation algorithm, where all the clustering schemes and their respective durations are calculated by the base station (BS) once at the network initialization. This provides the BS a global vision and enables it to reach better clustering schemes with adapted rounds' durations. It also eliminates costs proportional to periodic online re-clustering, using an energy prediction model. A new weight function is proposed to evaluate the chance of sensor-nodes to become CHs in a round. Experimental results demonstrate that the proposed protocol considerably prolong the network lifetime and the packet delivery ratio as compared to the LEACH-C protocol. The enhancement exceeds the double for the network lifetime and is more than ten times for packet delivery ratio.**

***Keywords*-Wireless Sensor Networks; Clustering Protocols; Energy Efficiency; Cluster-Head Election, Re-clustering Cost; Energy Prediction; Adaptive Round Time.**

I. Introduction

Clustering is a fundamental concept largely used in wireless sensor networks (WSN), where the latter is divided into clusters and one node is chosen as cluster-head (CH) in each cluster. This enables to achieve long lifespan in large-scale deployment. LEACH [1] is the first clustering-based protocol in WSN. Thenceforth, numerous solutions have been proposed. Those solutions can be classified according to several criteria, such as the technique of CHs selection and the methodology of clustering [2]. CHs are responsible for collecting data from their respective members, performing some data aggregation/fusion and forwarding the meaningful data to the base station (BS). The data traffic load tends to be concentrated at CHs, which makes CH selection an essential step in the clustering process. Optimal selection of CHs helps to further reducing energy consumption and extending the network lifetime.

Classification based on the methodology of clustering divides clustering protocols into two categories, distributed vs. centralized. Protocols of the first category have some advantages such self-adaptation, fast execution, and fault tolerance, but they may yield a non-optimal CHs choice [1]. In contrast, centralized clustering solutions offer better clustering quality since the BS (that runs the clustering computation algorithm) has a complete view of topology. However, this advantage comes at the cost of a high complexity process that repeats during the set-up phase in every round. In fact, for the sake of load-balancing, full-network-re-clustering is used in most protocols, where it is required that every node sends its energy level to the faraway BS at the beginning of each round. This has an important footprint in terms of energy consumption.

In this paper, we focus on the CHs selection process by proposing a new weight function to evaluate the chance of a node to become a CH in every round. The proposed function ensures a balanced energy consumption among all nodes, which in turn decreases the probability of early death of nodes. Further, we propose a new strategy for clustering schemes calculation based on an energy prediction mechanism that is used to determine the durations of rounds (when re-clustering should take place). The use of energy prediction also allows to prevent repetitive high-energy cluster reconstruction process while distributing the load evenly amongst all the nodes. Based on the proposed weight function and strategy, we propose a centralized one-step clustering protocol where the clustering schemes for the whole network lifetime are calculated once at the network set-up. The proposed solution targets periodic applications where nodes' traffic and energy consumption may be predicted beforehand, which eliminates the need for periodic reporting of nodes' energy states. To our knowledge, in all existing protocols, even those using estimation methods to reduce communication overhead, re-clustering schemes are determined progressively, i.e., the scheme for round $k+1$ is determined at the end of round k, without considering schemes of further rounds ($i > k+1$). This has the local optimum problem. The proposed protocol is the first that uses one-step off-line approach targeting global optimization. In our solution, all the clustering schemes and their respective times (durations) are calculated by the base station

(BS) once, at the initialization of the network. This provides the BS a global vision and enables it to calculate schemes with better quality than existing approaches.

The rest of this paper is organized as follows. Second section gives a brief overview of some existing clustering protocols. Third section presents the network and the energy models. Section four describes the proposed solution, and Section five discusses simulation results. Finally, Section draws conclusions.

II. Related Work

Numerous clustering solutions have been proposed for WSN. While every solution deals with specific clustering challenge and/or application requirements, extending the network lifespan is the main and common objective shared by all clustering protocols. CHs selection has very significant impact on the energy efficiency of the clustering protocol given the load concentration at those nodes. Heinzelman et al. proposed LEACH [1] that uses a probabilistic process to elect CHs. Each node chooses a random number between 0 and 1 and accordingly announces itself CH with a predefined probability. The authors also proposed LEACH-C protocol, which uses a more efficient CHs selection mechanism based on nodes residual energy. In LEACH-C, the BS calculates the average energy of the network, and only nodes with energy level higher than the average can compete for being CHs. Simulated annealing algorithm is then applied to choose CHs that minimize the total energy dissipated by non-CH-nodes. In [3], the authors proposed an enhanced version of LEACH-C by considering the distance between a node and the BS as a factor in evaluating the chance of the node to be a CH. HEED [4] selects CHs according to a hybrid function between node residual energy and node degree. DEACP [5] combines residual energy, distance among the CHs, distance between node and BS, and node degree in the definition of optimality for choosing well distributed CHs.

Another important factor that influences energy efficiency of clustering protocols, and thus the network lifetime, is the rounds' durations. Azim et al. [6] proposed a method to model rounds' durations in a way that prevents early death of CH nodes. The key idea is that rounds' durations should allow all nodes to act as CH once, and as non-CH in the other $(N/k - 1)$ rounds, where N is the number of nodes in the network and K is the number of CHs. Another solution is proposed in [7], where the round-time is calculated according to the number of nodes that are alive in the network. In addition, the authors propose that after the death of 50% of nodes, round-time is fixed. The motivation behind this is that the remaining energy of nodes at that time will be very low, almost $10 - 20\%$ of initial energy.

Periodic re-clustering of the whole networks has high cost in terms of energy and communication overhead. However, re-clustering is inevitable to balance the load on the network. Energy estimation based re-clustering protocols have then been proposed. Protocols of this category use prediction of nodes' remaining energy instead of directly sending energy level. For instance, in LEACH-CE [8] the set-up phase takes place once every eight rounds. In this phase, nodes send their energy information at the beginning of two successive rounds. Based on this information, the BS calculates the average of energy consumed by CHs and member nodes in one round. Consequently, the residual energy of each node can be estimated by subtracting the calculated consumption from the energy of the previous round. This energy estimation model is used in the upcoming eight rounds to eliminate the communication cost. A more evolved protocol is proposed in [9], where the set-up phase is executed in the first three rounds. In each one, all nodes send information about their remaining energy level and location to the BS. Using this information, the BS estimates the energy consumed by each node, and thus, its residual energy. The estimation is based on the hidden Markov model and optimized with the particle swam optimization algorithm. Contrary to LEACH-CE, set-up phase takes place at the first three rounds only. After that, no communication is expected between the nodes and the BS.

III. Network and Energy Models

A. Network Model

The network is represented by an undirected unit disk graph, $G = (V, E, B)$, where the set of vertices, V, represents the nodes, and E is the set of edges, $(u, v) \in E$ if both u and v are in the communication range of one another. B denotes the base station that is supposed to be energy unlimited node. The set of nodes to which a node, v_i, is directly connected (v_i's neighbors), is denoted by $\mathcal{N}_{v_i}$. This model represented by G reflects a network topology when nodes use a regular transmission range. However, nodes are supposed to be able to communicate directly with the BS by increasing their power when needed (to initially transmitting energy information or when selected as CHs). The network is supposed stationary after deployment, and nodes to be aware of their location. Traffic is supposed to be periodic, with all nodes generating data packets of the same length and at a fixed rate. Data aggregation is used by CHs, i.e., a CH compresses all the data received from its members into a single message and then transmits it to the BS. A clustering Scheme S is defined as a set of clusters, $S = \{(CH_i, m_1, m_2, \cdots m_j)\}$, that covers the network (the graph G). Every tuple $(CH_i, m_1, m_2, \cdots m_j)$ represents a cluster i with a cluster-head CH_i, and j members $(m_1, m_2, \cdots m_j)$. The set of all schemes for the whole network lifetime is then represented by a vector, C, of schemes.

B. Energy Model

We use an energy model based on the first order radio model [1]. The energy consumed in the transmission mode, E_{tx}, is dissipated to run the radio electronics and the power amplifier. To achieve an acceptable signal-to-noise ratio (SNR), amplification power depends upon the distance, d, between transmitter and receiver. If d is below a threshold, say d_0, then free space (fs) model is used; otherwise, multi-path fading (mp) propagation model is considered. Thus, to transmit an l bits packet over a distance, d, the radio expends the following amount of energy:

$$E_{tx}(l,d) = lE_{el} + l\epsilon_{fs}d^2, if(d <= d_0), \qquad (1)$$

$$E_{tx}(l,d) = lE_{el} + l\epsilon_{mp}d^4, if(d > d_0), \qquad (2)$$

where Eel is the energy required for running the transceiver electronic circuitry. l is the length of the transmitted packet. ϵ_{fs} and ϵ_{mp} are the amplification energy in the free space and multi-path fading models, respectively.

The crossover distance, d_0, is calculated as follows.

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \qquad (3)$$

In receiving mode, the energy, E_{rx}, is dissipated to run the radio electronics. To receive l bits, a node consumes the following amount of energy,

$$E_{rx}(l) = lE_{el}. \qquad (4)$$

Further, the approximate a energy consumed by node, v_i, in round, k, varies depending on whether the node is a member or CH in that round. Let frame be the unit of time. In each frame, a member node collects and sends data to its CH. CH receives data from its members, performs some data aggregation and transmits the processed data to the BS. Consequently, the energy dissipated in a time unit by a member node, v_m and a CH, v_h can be respectively calculated as follows.

$$E_{v_m} = E_{sen} + E_{tx}(l,d), \qquad (5)$$

$$E_{v_h} = E_{sen} + E_{rx}(l)\lambda_h + E_{agr} + E_{tx}(l,d), \qquad (6)$$

where E_{sen} and E_{agr} are the energy spent in sensing and data aggregation, respectively. λ_h is the number of members of CH v_h. The consumed energy of any node, v_i, in a round, r, is then given by,

$$E_{c_{v_i}}(r) = E_{v_i}t_r, \qquad (7)$$

where t_r is the round duration, i.e, time of the round in terms of number of frames. Accordingly, the expected remaining energy of node v_i can be obtained as follows.

$$E_{re_{v_i}}(r) = E_{re_{v_i}}(r-1) - E_{c_{v_i}}(r-1) \qquad (8)$$

IV. Proposed Solution

The proposed centralized clustering protocol– entitled OSC for One-Step Clustering– is presented in this section. The main objective of this protocol is the extension of the network lifetime. Although it performs in rounds, the clustering scheme is not calculated at the beginning of each round. Instead, all clustering schemes are calculated in one-step at the beginning of the first round, using an energy prediction mechanism. This has two advantages, i) first, the cost of periodic network re-clustering is eliminated while ensuring a proper distribution of the load among nodes, ii) second, rounds' durations are fixed in such a way to efficiently utilize the energy of nodes. Moreover, OSC proposes a new weight function to select the set of CHs that minimizes energy consumption of CHs, and non-CH nodes as well.

In OSC, each round consists of two phases, i) set-up phase, and ii) steady state phase. The set-up phase of the OSC differs from all existing clustering protocols in the fact that clustering schemes for the whole network lifetime are generated together in the set-up phase of the first round. To do this, nodes send their energy levels and positions to the BS at network initialization. BS then calculates all clustering schemes by selecting the optimal set of CHs and broadcasts a message (SCHEME) that holds the clustering scheme of the first round. At the reception of this message, every node knows whether it is selected as CH or to which cluster it belongs. From the set-up phase of the second round, only the BS broadcasts a (SCHEME) message containing the generated scheme for the incoming round that are already calculated. As for the steady state phase, it is similar to that of LEACH-C based protocols, where nodes perform periodical data gathering and send the collected data to their respective CHs at each frame. CHs aggregate the received data and relay to the BS.

OSC clustering approach is illustrated in Algorithm 1, to which we refer in the following description. The procedure $GENSCHEME$ is used to calculate a single clustering scheme. We propose a greedy approach based on a new weight function for CH selection (to be described later), where the BS selects the node with the lowest weight as the first CH (line 29) with its neighbors as members to form the first cluster (line 30). Next, the BS recalculates the weight of the remaining nodes (reduced graph by removing the selected CH and its members from G), and again the node with the lowest weight is selected as second CH and a second cluster is formed by this CH and its neighbors. The BS repeats this process until all nodes of the network are part of the clustering scheme (loop between lines 25 and 32). This way, the first clustering scheme (of the first round) is created.

For CHs selection, we propose a new weight function to evaluate the chance of nodes to be CHs in the incoming round. This function comprises node residual energy, as well as relevant topology features, i.e., node degree and distance between the node and the BS. Further, the average energy required to transmit a packet to that node from its neighbors is also considered. The weight function is given by the following equation,

$$\omega_{v_i} = \frac{E_{tx}(l, d_{v_i,BS}) EN_{v_i}}{\Delta_{v_i} E_{rx}(l) E_{re_{v_i}}}, \quad (9)$$

where Δ_i is node v_i degree in the reduced graph (number of its neighbors that are not assigned yet to a CH), and EN_{v_i} is the average energy consumed by the neighbors of node v_i (not assigned) to transmit a packet to that node. EN_{v_i} is calculated as follows,

$$EN_{v_i} = \frac{\sum\limits_{u_i \in \mathcal{N}_{v_i}} E_{tx}(l, d_{u_i,v_i})}{\Delta_{v_i}}. \quad (10)$$

The aim of this weight function, ω_{v_i}, is the selection of nodes with high remaining energy, minimum communication cost to reach BS, and those reducing the energy dissipated by their members. Those nodes are then chosen to act as CHs to burden the pressure of data transmission. It is clear from the formula that the likelihood for selecting a node as a CH is reflected by low values.

After the generation of the first clustering scheme at time t, the BS calculates the energy required to satisfy this scheme for one unit of timeframe (lines 6, 7). It uses E.q. 5 and E.q. 6 to calculate the energy that would be consumed by a member node and a CH, respectively, and E.q. 8 to estimate residual energy of each node after one frame. After that, and basing on the estimated residual energies, the BS generates a new clustering scheme for the unit of time $t+1$. The two clustering schemes are compared by the BS (line 10). Two schemes are considered similar they have a certain percentage of CHs and clusters in common. The optimal value of this percentage will be fixed empirically in the simulation phase. If the two schemes are considered similar, the BS keeps the old one for another unit of time and extends the current round duration (T_r) by one frame. Otherwise, the BS triggers a new round and keeps the new clustering scheme (lines 15 throughout 17). The aim in this phase is not to change the clustering of the network unless the change is considered to be advantageous in terms of energy efficiency. The BS repeats this process for the whole network lifetime (loop in lines 8 throughout 21). The network lifetime is defined by the exhaustion of a certain number of nodes that is application dependant. This makes high level of abstraction that captures any definition such as the first battery exhaustion of a node in the network ($\alpha = 1$), the exhaustion of a portion of nodes, e.g., half ($\alpha = N/2$), third ($\alpha = N/3$), all (N), etc.

Algorithm 1 OSC clustering algorithm

```
1: Input: G(V, E, B), α, E_re ∈ ℝ^N
2: Output: C_{j∈{1···K}} of schemes, r, T ∈ ℕ^K

3: r = 1, T = 𝟙, C = ∅
4: GENSCHEME(G, S)
5: C_r = S
6: Estimate E_c of all nodes using E.q. 5 and E.q. 6
7: Update E_re using E.q. 8
8: while (||{v_i ∈ V/E_re_{v_i} = 0}|| < α) do
9:     GenScheme(G, S')
10:    Compare S and S'
11:    if (S and S' are similar) then
12:        Maintain the scheme S for another frame
13:        T_r = T_r + 1
14:    else
15:        r = r + 1
16:        C_r = S'
17:        S = S'
18:    end if
19:    Estimate E_c of all nodes using E.q. 5 and E.q. 6
20:    Update E_re using E.q. 8
21: end while

22: procedure GENSCHEME (G(V, E, B), S)
23:    Input: G(V, E, B), Output: S

24:    V' = V; S = ∅
25:    while (V' != ∅) do
26:        for (each vertex v_i ∈ V') do
27:            calculate its weight ω_{v_i} using E.q. 9
28:        end for
29:        CH = arg min_{v_i∈V'} (ω_{v_i})
30:        S = S ∪ (CH, N_CH ∩ V')
31:        V' = V' \ (CH ∪ (N_CH ∩ V'))
32:    end while
33: end procedure
```

V. Simulation Study

To evaluate the OSC protocol, we have performed a simulation study with the network simulator NS3 [10]. Random topologies in an area of $100 * 100m^2$ have been generated with the GenSeN tool [11]. The parameters used in the simulation are summarized in Table 1. The performances of the proposed solution were compared to LEACH-C as a reference centralized clustering protocol. The results are presented hereafter.

Parameters	Values
Node Deployment Area	100m X 100m
Number of nodes	100, 150 and 200 nodes
BS Position	(50, 150)
Initial Energy (E0)	0.25 joule
Transmission Energy (Eelec)	50 nJoule/bit
Propagation Energy (free space Efs)	10 pJoule/bit/m^2
Propagation Energy (multi path Emp)	0.0013 pJoule/bit/m^4
Data Aggregation Energy (Eda)	5 pJoule/bit/signal
Threshold distance (d0)	87 meters
Packet Size	400 bits

Table I
SIMULATION PARAMETERS.

Figure 1 shows the obtained network lifetime for scenarios of 100, 150 and 200 nodes. The following definitions of the network lifetime are considered. i) FND, as the time

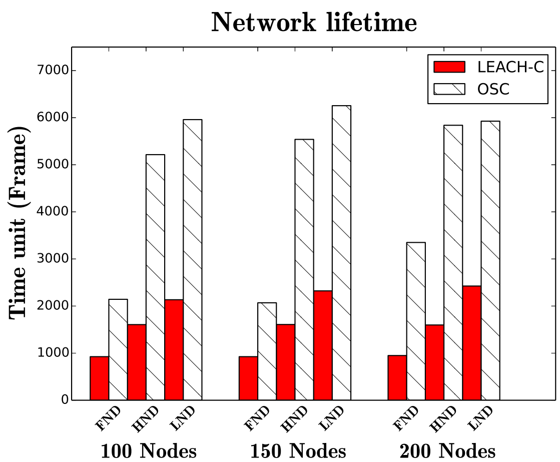

Figure 1. Network lifetime in homogeneous networks.

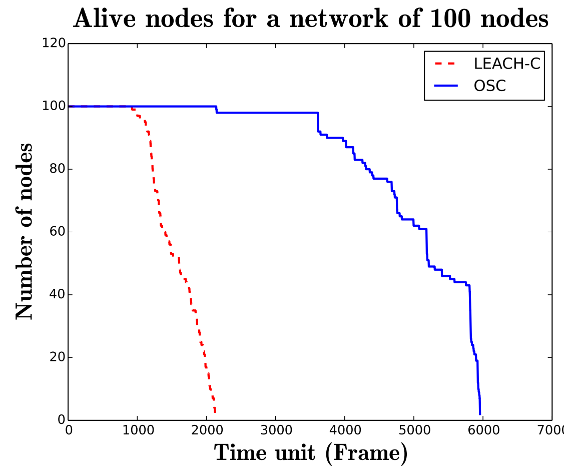

Figure 2. Alive nodes for a homogeneous network of 100 nodes.

interval from the start of network operation until the death of the first node, ii) the time interval until the death of half of the nodes (HND), and iii) the time interval until the last node death (LND). It is clear how OSC improves this metric. The enhancement over LEACH-C is on average more than doubling the network lifetime when considering FND and LND, and more than tripling the HND. This is due to the elimination of re-clustering and thus of direct transmissions in each round, which in turns saves a high amount of energy. The optimal choice of CHs and their members also plays an important role in minimizing the total energy consumed in the network. In fact, OSC chooses as CHs nodes in appropriate locations, with high residual energy, and minimizing the total energy dissipated by potential members. In addition, adopting dynamic durations of rounds contributes to further energy conservation.

In Figure 2, the number of surviving nodes is plotted vs. time. The figure clearly shows that OSC outperforms LEACH-C. Figures 3 and 4 confirm the obtained results for networks of 150 and 200 nodes, respectively. In LEACH-C, the number of surviving nodes start decreasing sharply after less than 2000 frames in all scenarios and all nodes become off batteries in less than 2500 frames, while OSC keeps almost all the nodes alive for more than 3000. The decrease is then relatively smoother than LEACH-C, and the time to the last node death is extended to around 6000 frames. The improvement with respect to this metric is due to the energy-efficient operations of the proposed solution, as explained previously.

In figure 5, the total number of data packets received by the BS is depicted. Since the proposed protocol extends the network lifetime and allows to keep nodes alive for longer time compared to LEACH-C, it is expected to have a higher rate in terms of the data packets received at the BS. The results of figure 5 confirm this and show a dramatic increase

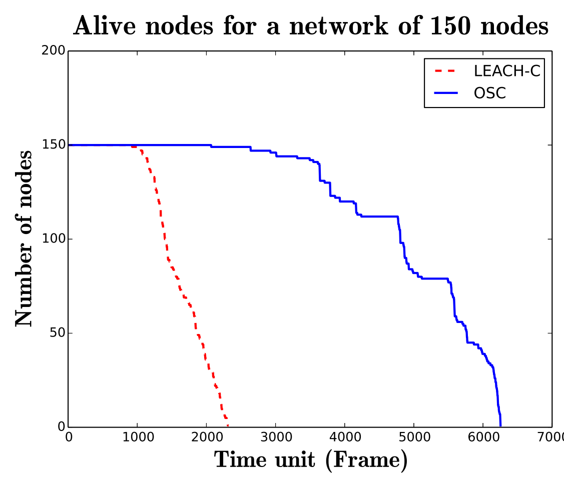

Figure 3. Alive nodes for a homogeneous network of 200 nodes.

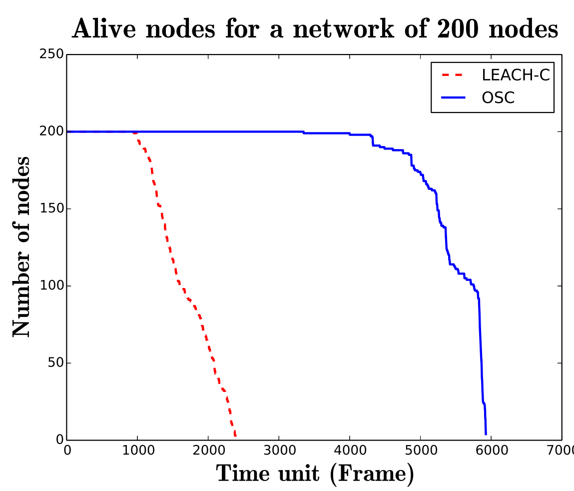

Figure 4. Alive nodes for a homogeneous network of 300 nodes.

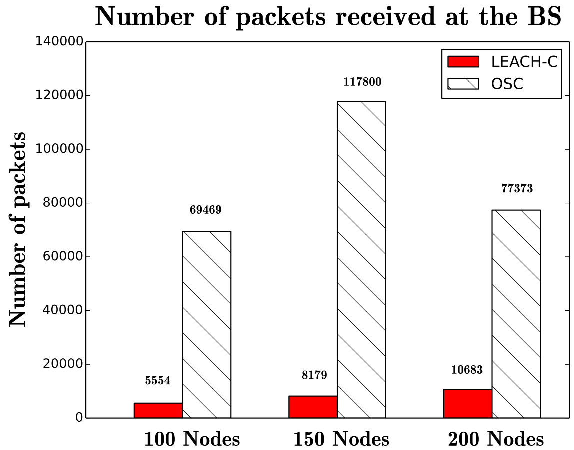

Figure 5. Packet received for homogeneous networks.

that exceeds ten times in the different network densities considered.

VI. Conclusion and Perspectives

The problem of energy efficient clustering in wireless sensor networks (WSNs) has been considered in this paper and new protocol has been proposed. The proposed protocol, called OSC (for One-Step Clustering), targets WSNs' applications with periodic traffic. It aims at reducing and balancing the overall network energy consumption by 1) selecting CHs that have high residual energy, require minimum energy to communicate with BS and minimize energy consumed by their members, 2) eliminating the cost of periodic full-network-re-clustering, which saves a high amount of energy, and 3) adapting the rounds' durations to the energy state of the nodes. One of the proper features of the proposed protocol compared to the state-of-the-art solutions is the computation of the whole clustering schemes at the initialization phase, with adaptation of every scheme's duration in accordance with the obtained clustering. A new weight function is also proposed to evaluate the chance of sensor-nodes to become CHs in every round.

The proposed protocol has been evaluated and compared to LEACH-C by simulation. The results confirm the efficiency of OSC and show clearly superiority as compared to the LEACH-C protocol. The time to the first and last node death are extended by more than double, while the time to half nodes death is enhanced by more than triple. Consequently, the number of received packets at the BS is increased by more than ten times.

Centralized clustering solutions based on energy estimation are known for the fault-intolerance problem. This is because no communication is expected between nodes and BS, except for transmitting data. That is, BS can elect a failed node as CH for the incoming round. Consequently, fault-tolerance is essential for effective deployment in real applications and represents a perspective of this work. Adapting the OSC protocol to the query-driven and the event-driven traffic based applications is also in our agenda.

References

[1] W. B. 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, Oct 2002.

[2] "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14/15, pp. 2826 – 2841, 2007.

[3] S. Shi, X. Liu, and X. Gu, "An energy-efficiency optimized leach-c for wireless sensor networks," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*, Aug 2012, pp. 487–492.

[4] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, Oct 2004.

[5] C. Gherbi, Z. Aliouat, and M. Benmohammed, "Distributed energy efficient adaptive clustering protocol with data gathering for large scale wireless sensor networks," in *Programming and Systems (ISPS), 2015 12th International Symposium on*, April 2015, pp. 1–7.

[6] A. Azim and M. M. Islam, "A dynamic round-time based fixed low energy adaptive clustering hierarchy for wireless sensor networks," in *Communications (MICC), 2009 IEEE 9th Malaysia International Conference on*, Dec 2009, pp. 922–926.

[7] V. Pal, G. Singh, and R. P. Yadav, "Network adaptive round-time clustering algorithm for wireless sensor networks," in *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, Aug 2013, pp. 1299–1302.

[8] J. m. Kim, H. k. Joo, S. s. Hong, W. h. Ahn, and H. b. Ryou, "An efficient clustering scheme through estimate in centralized hierarchical routing protocol," in *2006 International Conference on Hybrid Information Technology*, vol. 2, Nov 2006, pp. 145–152.

[9] R. Goudarzi, B. Jedari, and M. Sabaei, "An efficient clustering algorithm using evolutionary hmm in wireless sensor networks," in *Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on*, Dec 2010, pp. 403–409.

[10] G. F. Riley and T. R. Henderson, *The ns-3 Network Simulator*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 15–34.

[11] T. Camilo, J. S. Silva, A. Rodrigues, and F. Boavida, *GENSEN: A Topology Generator for Real Wireless Sensor Networks Deployment*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 436–445.