FLECHA: a Forecasting eLEction meCHAnism for semantic collectors sensor nodes

Mauricio M. Neto[1], Leonardo O. Moreira[2], Danielo G. Gomes[1]

[1]Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat/UFC)
Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI/UFC)

[2] Instituto Universidade Virtual (UFC Virtual)

Universidade Federal do Ceará (UFC)
Fortaleza – CE, Brasil

{mauriciomoreira,dgomes}@great.ufc.br, leoomoreira@ufc.br

Abstract. *Due to the resource constraints of the sensor nodes, energy provisioning in wireless sensor networks remains a challenging task, particularly in outdoor scenarios. Among the literature proposals to mitigate this problem, we highlight semantic clustering as a recent energy-efficient technique for prolonging the network lifetime. In semantic clustering, each cluster has a semantic leader (collector) which is periodically elected according to an energy-related criterion. However, since collectors' energy depletion is faster than the others members of their cluster, suitable election mechanisms are required to avoid the energy hole problem. Here we propose FLECHA, a mechanism based on the ARIMA model to predict semantic collectors elections with leader-node alternation. Our hypothesis is that by anticipating the best candidates to semantic leaders, we can improve the energy-saving at the node-level, and hence allow the network lifetime to be further extended.*

1. Introduction

Technological advances in microelectronics and wireless communications have enabled a fast development of the so-called Low-Power and Lossy Networks (LLN) [Remy 2015], including Wireless Sensor Networks (WSN), which are comprised of several sensing embedded nodes. Nowadays, these small-size nodes play a key role as a sensing and communication infrastructure for the smart-x applications [Somov and Giaffreda 2015, Mónton and Picone 2015], such as smart environment, smart metering, smart agriculture, and green smart cities [Gomes and Forster 2015]. Although there are interesting energy-harvesting solutions in outdoor and low power applications [Frohlich et al. 2015], most of the sensor nodes are battery-operated [Hermeto et al. 2013], making energy efficiency a critical goal.

,To achieve a higher energy efficiency related to data transmission, a commonly used technique in wireless sensor networks is clustering. In such approach, the whole set of network nodes is divided into partitions controlled by an elected leader node, namely Cluster Head (CH) [Deshpande and Bhagat Patil 2013]. To form the clusters, physical clustering commonly uses physical metrics such as Euclidean distance and residual energy [Jannu and Jana 2014]. The proximity of the nodes is one of the main parameters of clustering in a physical clustering. Semantic clustering, instead, uses the similarity of data collected by nodes, i.e., when certain nodes detect the same kind of data, the semantic cluster is created [Hermeto et al. 2013, Rocha et al. 2012]. The semantic clustering also uses the concept of leader (semantic collector), analogous to the CH in a physical cluster.

Concerning the differences between physical and semantic clustering, we highlight that a semantic cluster is independent of physical metrics, such as distance, so that it can be created with separate physical cluster members. It is important to realize that a physical leader can also become a Semantic Collector (SC). However, in case we have two semantic clusters, their collectors cannot be the same node. Given that the semantic clusters are created using data similarity, it is not possible for a same semantic node to join two or more groups. In both physical and semantic clustering, each leader is responsible for receiving data from adjacent semantic nodes (semantic neighbors), summarizing and sending them to the sink node via multihop communication [Avril et al. 2014]. Although energy efficiency in the semantic clustering is better than in the physical clustering [Rocha et al. 2012], the nodes elected as leader also tend to lose a large amount of energy, which leads them to die prematurely. To avoid the energy hole problem, algorithms based on leader-node alternation are required [Jannu and Jana 2014].

Here we present **FLECHA**, a **F**orecasting e**LE**ction me**CHA**nism for semantic collectors in wireless sensor networks. FLECHA uses ARIMA (Auto-regressive Integrated Moving Average) models as the decision mechanism to improve the election of new semantic leaders. ARIMA is considered a general class of models for forecasting a time series. Besides, it has performed well on short-term prediction windows [Moreira et al. 2014, Santos et al. 2013], which can be suitable to avoid temporal disruptions in WSN and other short-term communication failures.

2. Related Work

This section is divided in two subsections that surrounds prior works related to FLECHA, which involve clustering and prediction in WSN.

2.1. Clustering and Prediction in Wireless Sensor Network

Genetic algorithms can be used to select cluster heads in a centralized cluster and can also prolong the network lifetime compared to other clustering protocols such as LEACH [Heinzelman et al. 2002] and LEACH-C [Pal et al. 2015]. The work proposed by Pal et al. (2015) resulted in an improvement of the load balance in relation to a traditional clustering. However, the centralized network approach generates an overhead, causing a greater increase in energy expenditure in the nodes of the network. Our approach is decentralized, resulting in an autonomy of sensor nodes and thus eliminating the network overhead.

The residual energy and distance among the nodes can be used as parameters to elect cluster leaders in a heterogeneous WSN, as DTRE-SEP [Hassan et al. 2015]. The work proposed by Hassan et al. (2015) presented an improvement of the network lifetime in unstable regions compared to clustering algorithms as LEACH. However, such protocols do not have a good performance in networks with random distributions of nodes [Hassan et al. 2015]. FLECHA has autonomic characteristics as self-organization, self-configuration and self-adaptation, having a good performance in random distribution of nodes.

Algorithms based on static and dynamic hierarchical levels of clusters, together with filters to predicts targets positions of the nodes, are used to improve WSNs lifetime, such as proposed by PRATIQUE-a [Souza et al. 2015]. The forecast methods could be utilized to obtain the set of nodes that will detect the next event, reducing the overhead of messages in the network [Souza et al. 2015]. However, the PRACTIQUE-a does not use metrics like residual energy as a parameter to choose cluster leaders, which could generate an improvement in the WSNs lifetime.

2.2. Wireless Sensor Network and Semantic Clustering

The semantic clustering is used for event monitoring in a WSN and its cluster is formed by the similarity of the data captured by the nodes, which elect a semantic leader (collector) [Hermeto et al. 2013, Rocha et al. 2012]. The semantic clustering has better energy efficiency than other classical clustering methods [Rocha et al. 2012].

It is possible to improve the lifetime of the semantic clustering using fuzzy logic to elect the semantic collector [Hermeto et al. 2013]. Using metrics such as residual energy and Euclidean distance it is feasible to have an alternation of leaders to decrease the overhead in the collector (leader) [Hermeto et al. 2013]. FLECHA uses a predictive method to anticipate the election of a new semantic collector through this proactive approach, making it is possible to improve the lifetime of the WSN.

It is possible improve the energy efficiency of a WSN with semantic clustering by introducing a mechanism of decentralized sensor nodes [Rocha et al. 2016]. Fully-decentralized mechanisms tend to have better energy performance than protocols partially decentralized by the reducing the amount of messages and the number of hops in the network [Rocha et al. 2016].

Table 1. Summary of related work.

Reference	Cluster	Self-*	Leader election technique	Leader election parameters
FLECHA (proposal)	Semantic	Organization, configuration and adaptation	ARIMA (prediction)	Residual energy, Euclidean distance
DSENSE [Rocha et al. 2016]	Semantic	Organization, configuration and adaptation	N/A	Fuzzification Message
Hermeto et al. (2013)	Semantic	Organization, configuration and adaptation	Fuzzy Logic	Residual energy
SEMANTK [Rocha et al. 2012]	Semantic	Organization, configuration and adaptation	N/A	Amount of Neighbors
PRATIQUE-a [Souza et al. 2015]	Physical	N/A	N/A	N/A
Pal et al. (2015)	Physical	N/A	Genetic Algorithms	Residual energy
DTRE-SEP [Hassan et al. 2015]	Physical	N/A	N/A	Residual energy, Euclidean distance

Table 1 shows that most related works use the residual energy as a key parameter to (re)elect leaders. SEMANTK and DSENSE use semantic clustering to improve the network lifetime, but the decision to choose the collector is reactive. It is possible to further decrease the energy cost of the WSN using a predictive selection of semantic collectors. FLECHA employs ARIMA to generate predictive curves having residual energy and Euclidean distance as input parameters and, through the trends of these curves, to elect the new semantic leaders of the cluster, avoiding the excessive energy cost.

3. FLECHA: a Forecasting eLEction meCHAnism for semantic collectors sensor nodes

FLECHA is a mechanism to elect leaders (collectors) in a WSN based on semantic clustering which metrics are the residual energy and Euclidean distance among nodes. FLECHA extends SEMANTK [Rocha et al. 2012] by using the same oriented events mechanism. Figure 1 shows the FLECHA UML diagram where we can see three modules or phases (separated by colors): (i) data collection and analysis of sensor nodes (yellow); (ii) the detection of semantic neighbors in the physical cluster (orange) and (iii) the predictive election of semantic collectors mechanism (blue). All the FLECHA processes occur locally in each node. In the first phase, each sensor node will collect metrics (e.g., temperature, humidity) based on the application of interest. Moving on the next phase, a new event is expected (e.g., “frequency values higher than *X*”). In case of the event does not occur, FLECHA comes back to the initial data collection phase. However, if a new event is detected, the detection of semantic neighbor phase on the physical cluster is triggered. This phase is responsible for establishing the relationship between the semantic neighbors based on data similarity [Rocha et al. 2016].

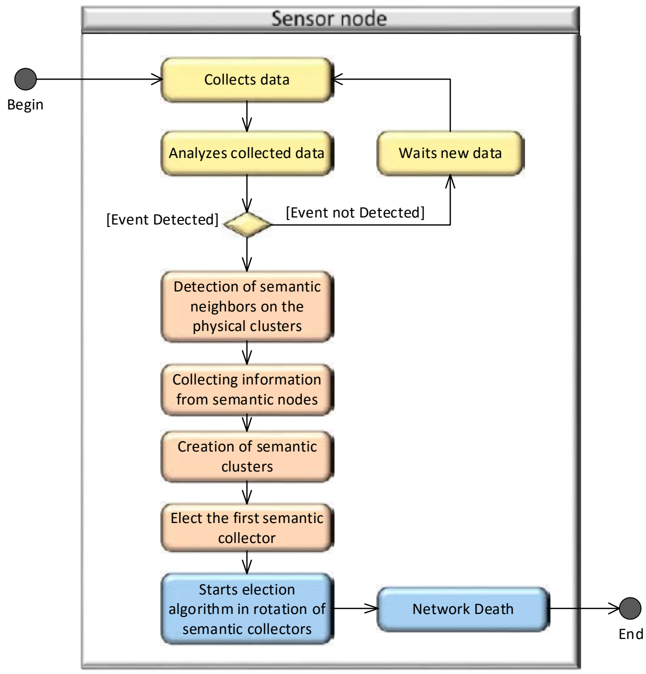

Figure 1. FLECHA UML diagram.

The election of the first semantic collector takes into account the Euclidean distance metric. Despite the main metric considered in FLECHA is the residual energy, it is initially supposed that all nodes of the semantic neighborhood have approximately the same level of residual energy. Consequently, we need to use other metric different from residual energy to elect the first semantic collector. In this work, the node with more semantic neighbors will be the first collector.

The subsequent phase is the predictive election for the next semantic collectors. It is common the cluster leader tends to die before the other nodes of clusters. The semantic collector (leader of the semantic cluster) tends to have a greater energy expenditure than the other sensor nodes of the semantic cluster that only capture the event data.

In our proposal, the election of the next semantic collector occurs after 10% of power depletion of the current collector, which is a reactive approach. To adopt a proactive one, we use ARIMA prediction method to generate the trend curve of residual energy decrease for the collector node. The election of the new semantic collector starts when the trend of the predictive curve is close to 10% of power drop in a current collector (see Algorithm 1). It can be seen by the algorithm 1 that the prediction curve *y[t + n]* is fed by the residual energy data *x[t]*, where *t* is the current instant of energy captured and *n* is the future moment generated by the predictive method.

```
Data: collector election
Result: elect a new semantic collector
initialization (predictiveCurve);
collected data (residualEnergy, euclideanDistance);
x[t] ← residualEnergy;
y[t] ← predictiveCurve;
z ← euclideanDistance;
if |y[t + n]| >= |0.1.x[t]| then
    collects all the energy and distance data of the semantic cluster nodes;
    elect the node with larger x;
    if nodes tied == TRUE then
        choose the node with more semantic neighbors using z;
    end
else
    go back to the beginning;
end
```

Algorithm 1: Semantic Collector Election (pseudo-code).

During a new collector election, it is possible that two or more nodes have the same level of residual energy. In this case, the node with more semantic neighbors will be elected as a collector. Finally, when all the nodes of the semantic cluster reach a residual energy value below 10% or the first collector die (First Node Death indicator, see Section 4.2), the WSN will be considered dead.

By using ARIMA, our aim is to elect semantics collectors with better conditions to perform their duties. Thus, we seek to extend the network lifespan by choosing the best semantic collector for the next status of the network. ARIMA was chosen because it performs better with small prediction windows [Santos et al. 2013] [Moreira et al. 2014]. This model can be used to predict certain events like time series showing trends, correlations, and seasonal variations.

4. Material and Methods

The FLECHA mechanism was implemented in C using the Contiki operating system [Dunkels et al. 2004]. The simulations were performed on Cooja [Osterlind et al. 2006], which is used on the communication standard of IEEE 802.15.4 and 250 kbps transmission rate. Cooja also supports various sensor platforms; here we used the MicaZ platform.

4.1. Energy Model

We used the MicaZ energy model as follows [Jurdak et al. 2008]:

$$E_t = P_{send} * P_{length} * TB * I_t * V \quad (1)$$

$$E_r = P_{receive} * P_{length} * TB * I_r * V, \quad (2)$$

where E_t and E_r are, respectively, the energy costs of transmission and packet reception (mJ). P_{send} and $P_{receive}$ represent the number of packets sent and received. P_{length} is the packet size, and TB is the radio response time. I_t and I_r are the respective values of the node radio electric current in the modes of transmission and reception. Lastly, V represents the voltage constant provided by the MicaZ datasheet.

4.2. Evaluation Scenarios

We used the scenarios and values from Hermeto et al. (2013) (Figure 2) but adding more sensor nodes in different situations. We simulated an application for structural health monitoring (SHM) domain as a case study. With a scenario of a five floors building, all physical clusters are deterministic.

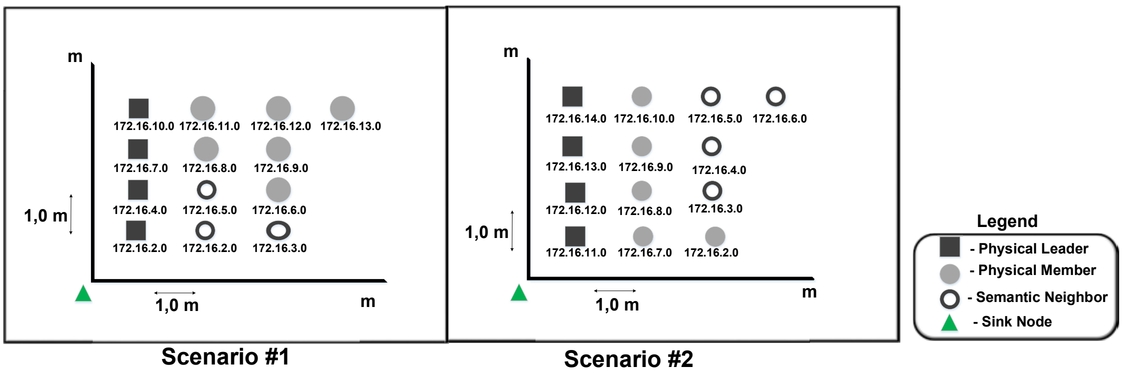

Figure 2. Simulated scenarios (adapted from [Hermeto et al. 2013]).

In these scenarios, the events are triggered simulating damage in the physical structure, which will cause a change of state (shown by modal frequencies, Table 2). When the network starts to operate, all the cluster's nodes know their respective physical leaders as well the physical leaders know where adjacent leaders are. The sensor nodes transmission time ranged from 500 to 1300 ms and each round lasted 15 min.

At the simulation beginning, all physical nodes are healthy. During the first ten seconds, the physical clusters nodes send messages with the modal frequencies shown in

Table 2. Afterwards, we forced FLECHA to change all nodes frequencies to "Damage 1" for scenario #1 and "Damage 2" to scenario #2, thus causing damage on the 2nd floor (please see Figure 2).

Table 2. Modal Frequencies.

Structural Condition	Mode 1 (Hz)	Mode 2 (Hz)	Mode 3 (Hz)	Mode 4 (Hz)	Mode 5 (Hz)
Damage 1	2.34	7.52	11.62	14.45	17.48
Damage 2	2.44	6.93	11.91	15.14	16.99
Healthy	2.54	7.52	12.01	15.53	17.77

Based on the Euclidean distance, we first identified the nodes closer to the damage. Then, they became semantic neighbors by distribution of the nodes weights. In scenario #1, the IP nodes are 172.16.6.0, 172.16.7.0 and 172.16.8.0. Similarly, in the scenario #2, we used IPs 172.16.9.0, 172.16.11.0, 172.16.13.0 and 172.16.14.0. The event takes place when the frequencies of the nodes are changed, thus generating semantic clustering. When the semantic clustering occurs, the physical leaders know whether there are semantics nodes in their clusters or not. In our simulations, we extended the scenario used in [Hermeto et al. 2013] by doubling the number of nodes.

Since our aim is to reduce the premature loss of semantic collectors, we used the First Node Death indicator [Pal et al. 2015, Dietrich and Dressler 2009]. Our predictive approach was compared to SEMANTK's election algorithm, which selection criteria for the new semantic leader is the amount of semantic neighbors that exist on each physical cluster, and also to the Hermeto et al.(2013), who used fuzzy logic to elect the leader through the residual energy and distance among nodes.

5. Results and Discussion

Due to their multiple functions, leader nodes have higher energy costs than their neighbors, so that leaders tend to die faster. Thus, we considered the initial energy of all nodes as 5 J. Figure 3 shows the comparison of the semantic collectors lifetime in the scenario #1 between the approaches (a) SEMANTK (b) Hermeto et al. (2013), and (c) FLECHA. Likewise, Figure 4 presents a comparison of lifetime of the semantic collectors in the scenario #2 between the approaches (a) SEMANTK, (b) Hermeto et al. (2013), and (c) FLECHA. Both SEMANTK, Hermeto, and FLECHA algorithms were repeated 20 times for each experiment with 90% confidence intervals. Note that the vertical bars are not shown in figures where they are not visually significant.

Notes that Hermeto's algorithm took about 350 min to lose the first semantic collector. This result illustrates the gain in survival of about 45% of Hermeto et al. (2013) algorithm with regard to SEMANTK.

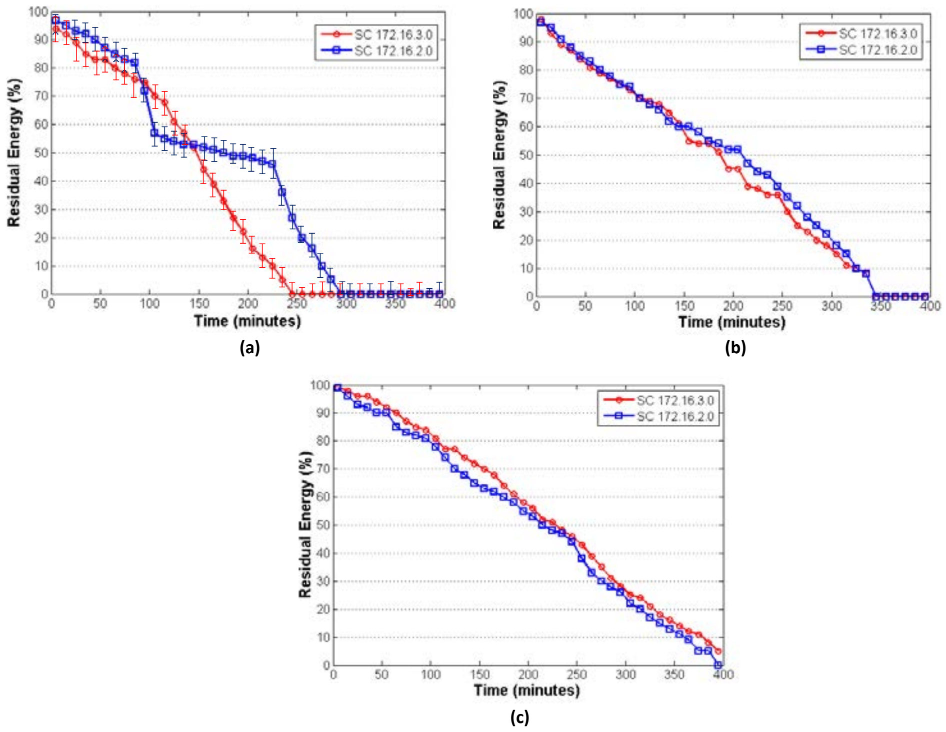

Figure 3. Comparison between approaches in scenario #1.

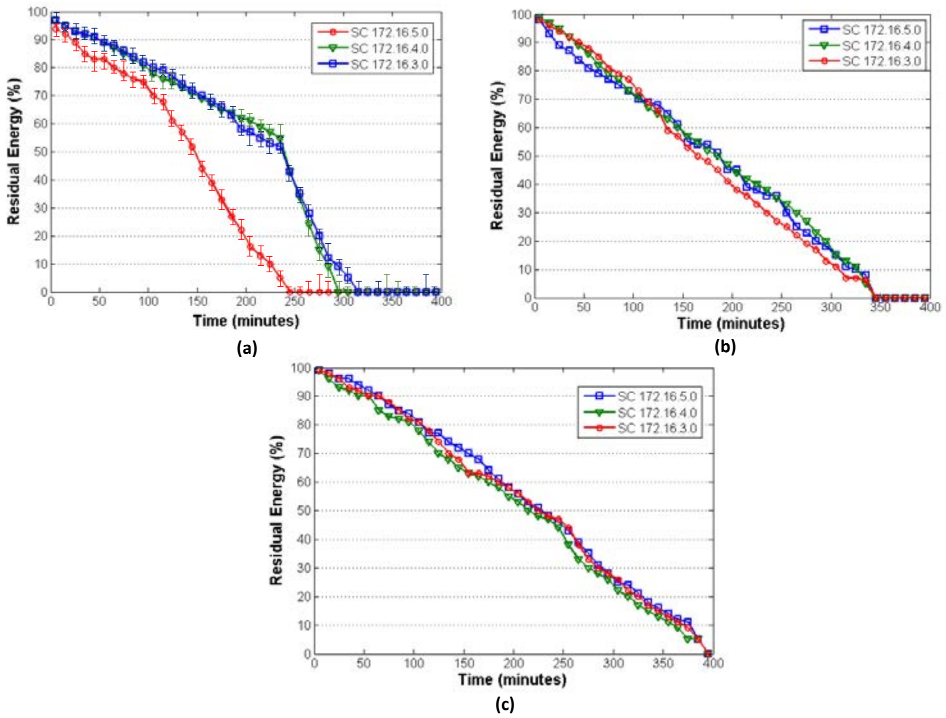

Figure 4. Comparison between approaches in scenario #2.

Since the time when the first semantic node die was around 400 min in the FLECHA approach, it brought a gain of 14% survival compared to the Hermeto approach and of 60% compared to SEMANTK algorithm. It can also be seen from Figures 3 and 4 that the first node was chosen by ARIMA, approximately, during the first 15 minutes of simulation of two scenarios. The last node was elected when it achieves 90% loss of energy. Using FLECHA, when the residual energy nears 10% of the energy loss in the initial state, ARIMA predicts time which will reach this loss. Therefore, the election is made in a predictive way before reaching 10% loss (on average 9%).

Table 3 illustrates the number of nodes chosen in scenarios #1 and #2. We can see in that the SC 172.16.3.0 was elected more times in Hermeto algorithm and FLECHA. We can also note that SC 172.16.5.0 has a better performance in SEMANTK and FLECHA when compared to Hermeto proposal. However, the other nodes elected as collectors (SC 172.16.4.0 and SC 172.16.3.0) present better results through FLECHA and Hermeto algorithm because of the proximity to the sink node.

Table 3. Quantity of nodes elected

Scenario	IP Address	SEMANTK	Hermeto et al. (2013)	FLECHA
#1	172.16.2.0	16	11	12
#1	172.16.3.0	0	12	14
#2	172.16.3.0	0	10	12
#2	172.16.4.0	1	9	9
#2	172.16.5.0	15	8	14

Afterwards, we doubled the number of nodes in scenario #1, which means it had two more candidates to be elected as semantic collectors (SC 172.16.4.0 and SC 172.16.5.0).

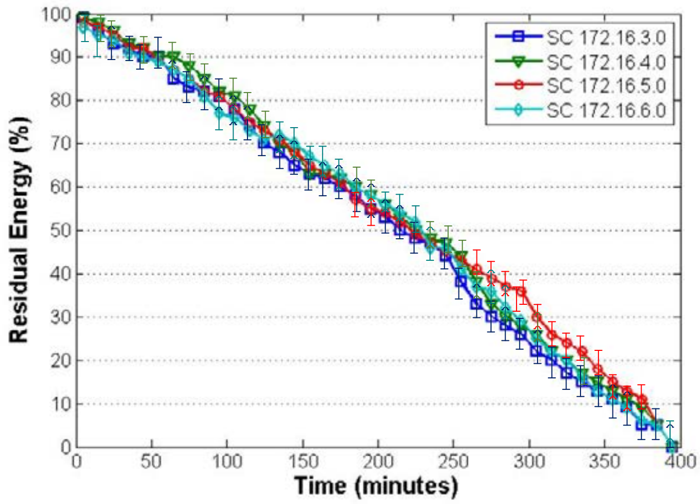

Figure 5. FLECHA execution in scenario #1 with twice nodes.

Figure 5 shows the scenario #1 with twice elected nodes as collectors by FLECHA. When we doubled the nodes number, the network lifetime (400 min) becomes similar to that in Figure 4(c). This is because FLECHA uses ARIMA as predictor to avoid energy waste on the collector node, which causes a new election per collector. As a result, FLECHA allows a gradual network energy depletion and, consequently, increases the network lifetime.

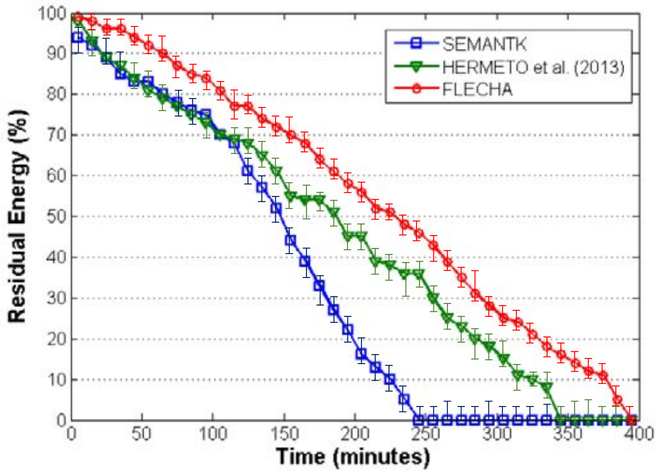

Figure 6. Approaches comparisons in scenario #1

In Figure 6, we see the comparison between SEMANTK, Hermeto et al. (2013) and the FLECHA for the scenario #1. There we note FLECHA brought a gain of 60% in network lifetime when compared to the SEMANTK and 14% compared to Hermeto's algorithm.

In this paper, the leader-node alternation is independent of how far the candidates are from the collectors. This is due to the load balancing of the input variables weights. It should be noted that when an event of interest is detected by a large amount of neighboring clusters, it will lead to an intense exchange of messages among the candidates for collectors until there is a consensus for a new leader election.

6. Conclusion

Here we propose FLECHA, an approach to elect WSN semantic collectors using ARIMA models to forecast the best candidates. FLECHA uses the residual energy of the nodes and the distance between semantic collectors. According to our results, FLECHA improved the network lifetime by 60% compared to SEMANTK [Rocha et al. 2012] and by 14% in relation to Hermeto's proposal [Hermeto et al. 2013]. These results provide insight into the suitability of forecasting algorithms for use in WSNs clustering in such a way that the constrained sensor resources are satisfied.

As future perspectives, we intend to extend and improve FLECHA to make it flexible for mid- and long-term prediction (e.g. using machine learning algorithms) and considering other metrics, such as the most central node, processing time, and memory footprint.

References

Avril, F., Bernard, T., and Bui, A. (2014). Efficient communication scheduling in clustered wsn. In *Computers and Communication (ISCC), 2014 IEEE Symposium on*, pages 1–6.

Deshpande, V. and Bhagat Patil, A. (2013). Energy efficient clustering in wireless sensor network using cluster of cluster heads. In *Wireless and Optical Communications Networks (WOCN), 2013 Tenth International Conference on*, pages 1–5.

Dietrich, I. and Dressler, F. (2009). On the lifetime of wireless sensor networks. *ACM Trans. Sen. Netw.*, 5(1):5:1–5:39.

Dunkels, A., Gronvall, B., and Voigt, T. (2004). Contiki - a lightweight and flexible operating system for tiny networked sensors. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, pages 455–462.

Frohlich, A. A., Bezerra, E. A., and Slongo, L. K. (2015). Experimental analysis of solar energy harvesting circuits efficiency for low power applications. *Computers and Electrical Engineering*, 45:143 – 154.

Gomes, D. G. and Forster, A. (2015). Introduction to the special issue on green engineering: Towards sustainable smart cities. *Computers and Electrical Engineering*, 45:141 – 142.

Hassan, S., Nisar, M., and Jiang, H. (2015). Dtre-sep: A direct transmission and residual energy based stable election protocol for clustering techniques in hwsn. In *Communication Software and Networks (ICCSN), 2015 IEEE International Conference on*, pages 266–271.

Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4):660–670.

Hermeto, R. T., Kridi, D. S., Rocha, A. R., and Gomes, D. G. (2013). A distributed algorithm for semantic collectors election in wireless sensors networks. *Journal of Applied Computing Research*, 3(12):1–10.

Jannu, S. and Jana, P. (2014). Energy efficient grid based clustering and routing algorithms for wireless sensor networks. In *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pages 63–68.

Jurdak, R., Ruzzelli, A., and O'Hare, G. (2008). Adaptive radio modes in sensor networks: How deep to sleep? In *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on*, pages 386–394.

Mónton, M. and Picone, M. (2015). An open-source cloud architecture for big stream iot applications. *Interoperability and Open-Source Solutions for the Internet of Things: International Workshop, FP7 OpenIoT Project, Held in Conjunction with SoftCOM 2014, Split, Croatia, September 18, 2014, Invited Papers*, 9001:73.

Moreira, L. O., Farias, V. A. E., Sousa, F. R. C., Santos, G. A. C., Maia, J. G. R., and Machado, J. C. (2014). Towards improvements on the quality of service for multi-tenant rdbms in the cloud. In *Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference on*, pages 162–169.

Osterlind, F., Dunkels, A., Eriksson, J., Finne, N., and Voigt, T. (2006). Cross-level sensor network simulation with cooja. In *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, pages 641–648.

Pal, V., Yogita, Singh, G., and Yadav, R. (2015). Cluster head selection optimization based on genetic algorithm to prolong lifetime of wireless sensor networks. *Procedia Computer Science*, 57:1417 – 1423. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

Remy, L. (2015). Smart gateway for low-power lossy networks. In *Proceedings of the 2015 on MobiSys PhD Forum*, PhDForum '15, pages 13–14, New York, NY, USA. ACM.

Rocha, A. R., Delicato, F. C., Pirmez, L., Gomes, D. G., and de Souza, J. N. (2016). A fully-decentralized semantic mechanism for autonomous wireless sensor nodes. *Journal of Network and Computer Applications*, 61:142 – 160.

Rocha, A. R., Pirmez, L., Delicato, F. C., Érico Lemos, Santos, I., Gomes, D. G., and de Souza, J. N. (2012). Wsns clustering based on semantic neighborhood relationships. *Computer Networks*, 56(5):1627 – 1645.

Santos, G. A. C., Maia, J. G. R., Moreira, L. O., Sousa, F. R. C., and Machado, J. C. (2013). Scale-space filtering for workload analysis and forecast. In *2013 IEEE Sixth International Conference on Cloud Computing*, pages 677–684.

Somov, A. and Giaffreda, R. (2015). Powering iot devices: Technologies and opportunities. *IEEE Internet of Things Newslettter*.

Souza, E. L., Pazzi, R. W., and Nakamura, E. F. (2015). A prediction-based clustering algorithm for tracking targets in quantized areas for wireless sensor networks. *Wirel. Netw.*, 21(7):2263–2278.