Energy- Efficient Cluster-Based Protocol using An Adaptive Data Aggregative Window Function (ADAWF) for Wireless Sensor Networks

Somasekhar Kandukuri, Jean-Mickaël Lebreton, Nour Murad, Richard Lorion, Jean-Daniel Lan Sun Luk

Energy-Efficient Cluster-Based Protocol using An Adaptive Data Aggregative Window Function (A-DAWF) for Wireless Sensor Networks

Somasekhar Kandukuri, Jean Lebreton, Nour Murad, Richard Lorion, and Jean-Daniel Lan-Sun-Luk
Laboratory of Energy, Electronics and Process (LE^2P), University de La Reunion, 15 avenue Rene Cassin CS 92003
97744 Saint Denis CEDEX 9, Reunion - France, Email: somasekhar.kandukuri@univ-reunion.fr

***Abstract*—We present an adaptive data aggregative window function (A-DAWF) for a distributed sensor network model in which nodes store data in their attribute window functions, and provide non-correlated data towards the base station (BS). Unlike previous works, namely data collection or data gathering management systems, we propose a novel approach that aims to process temporal redundant techniques in sensor nodes as well as providing spatial redundant filtration methods in cluster-head (CH) nodes. In this regard, preliminary results show that A-DAWF can suppress up to 90% of temporal redundant data among the considered sensor nodes by an optimal threshold of the window sizes, and their spatial correlations in CH node by a maximum error threshold compared to either periodic or a continuous data transmission system.**

I. Introduction

In WSNs, sensor nodes lifetime always rely on each other. In most of the application scenarios, sensor nodes (SNs) are used as battery powered devices which have limited storages and processing capabilities. Periodically, sensors can spend a lot of energy to transmit or receive the sensor readings which shortens their lifetime as well as their network lifetime. Since data aggregation or data collection has grown as one of the promising area in energy-efficient WSNs for maximizing the network lifetime, and several well-known data aggregation techniques have been studied in the literature [1–8], however they are limited to their specificity. Especially, earlier literature works are designed based on the homogeneous application scenarios than heterogeneous, since many deployment scenarios proved that nodes have their own specific application tasks [9] rather than homogeneous tasks which are rarely used in practice. Having a better aggregation mechanism not only reduces redundant transmissions, it also saves a huge amount of receiver's energy of all other three modules such as radio receiving or listening, computation, and processing procedures of receiver nodes. Hence, it is necessary to use an effective data redundancy technique or simple prediction methods in order to avoid the redundant data transmissions to ensure the reliability in network applications.

In this paper, we primarily designed our proposition based on the homogeneous application scenarios, and then heterogeneous. In this regard, we present two sorts of data filtration, one performs in sensor nodes for finding temporal data redundancies (TDRs) using both relative variation (RV) and aggregative window functions, and the another one uses in CH nodes for exploiting spatial data redundancies (SDRs) using both RV and A-DAWF. This paper is presented as work in progress, and the designed implementations of our work are discussed in section III. In section IV, we show the preliminary results to demonstrate that the proposed mechanism can suppress a huge amount of data transmissions in both time and space.

978-1-5090-2185-7/16/$31.00

II. Related Work

Energy-efficienct strategies are a widely studied area in WSNs, and a taxonomy of various approaches are presented in the literature [1].

TiNA [3] used a clause condition for specifying the differed ranges, if the differed range is greater than the specified range between any two values, then the differed result can be reported, otherwise ignored. TiNA is more related to our work, as we also used the RV function to find TDRs between every two window stored phenomena among the individual nodes. CAG [2] proposed a cluster-based technique which reports only spatial correlations of cluster nodes by a CH to the BS, and ignores the individual nodes temporal data. The authors in [7] proposed another cluster-based method like CAG to build a predictive model on CH nodes instead of individual sensor nodes and let complete computational burden on header nodes itself. In the contrary, the authors in [4] proposed a TinyDB and Cougar data collection models for maintaining node as a small database query engine to the BS, and the other recent works in [5, 10] enables a data storage with coffee file system (CFS), and a dynamic Antelope database system in every sensor node.

Unlike previous works, our proposal avoids the consideration of building either rich spatial compression-based or temporal prediction-based mathematical models, and designed a simple data redundancy algorithm based on the Intel datasets. The basic window principle of this work has been approached from one of our earlier research works [11]. In our considerations, normal nodes sense the environmental phenomena by using A-DAWF mechanism for exploiting TDRs in every node, and SDRs in CH nodes as well.

III. A-DAWF Preliminaries

In this paper, we exploit both temporal and spatial correlations through the attribute window functions in order to suppress the redundant data information at a fixed window time intervals r and R. We consider a cluster-based sensor network with n SNs, which continuously forward the uncorrelated set of data attributes to the CH, $TDR(t) = (p_1, p_2, ..., p_M)$ generates the sensed physical phenomena at different time instances t, and s CHs as super nodes, which receive spatial data messages of nodes at given time instances τ, $SDR(\tau) = (q_1, q_2, ..., q_N)$ and their own sensor readings. The attribute detection nature of environmental phenomena p_M, may be the attributes being sensed by nodes as temperature or humidity or may be the result of any application phenomena. If the sensor monitoring attributes are continuous, we consider A-DAWF mechanism to monitor the redundant data.

A. A-DAWF Implementation

A-DAWF is an aggregative window function which can reduce the redundant data transmissions for individual sensor nodes in both time and space. In this approach, we present two sorts of scenarios to do the filtration in both sensor nodes and CHs. In this work, we assume that the network model is single-hop clustering as shown in the figure 1, and the sensor nodes are purely distributed, which can compute and process the obtained environmental data at different time instances t by using A-DAWF mechanism. Since computation is the second highest energy consumer after the communication, although for computing the simple prediction-based models does not require much energy to make a comparison between the sensed readings through the window function. Thus, for suppressing

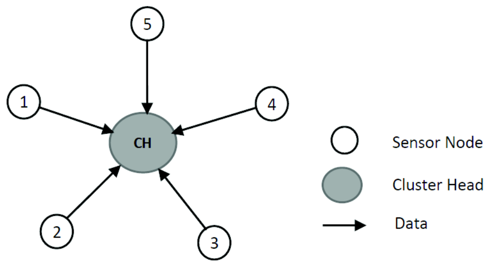

Fig. 1: Example of a cluster-star topology, show the nature of communication between sensor nodes and CH.

TDRs, the following constraints must be satisfied in sensor nodes.

$$TR_{count} = \begin{cases} if \frac{|p_{i+1}-p_i|}{p_i} > \eta, & TR_{count}= 0 \\ otherwise, & TR_{count}= 1 \end{cases} \quad (1)$$

$$\mu_K = \frac{\sum_{i=1}^{M} p_{i+KM}}{M} \quad (2)$$

According to the first scenario, the RV function of Eq.(1), where TR_{count} represents the temporal count of RV. If RV is greater than the threshold η then the TR_{count} will be 0, otherwise TR_{count} will be reported as 1. Therefore, we set the threshold η value as 0.05 and it can also be varied as per the physical phenomena or different application physical quantities. In the second scenario, A-DAWF evaluates the stored readings by Eq.(2) where μ_K is the mean value of window at K times which starts from zero. And i is an index of environmental property being sensed by sensor node as either temperature or humidity at given t time instances, and r is the round time interval of w_M. And, M is a window size that can either be fixed or vary based on the nodes computational resources or flash memory. During every w_M of a sensor node, it can compare its current value p_{i+1} with the previous value of p_i through the window stored readings by Eq.(1). In general applications, most of the obtained environmental data are either redundant or correlated. We assume that the observed nodes data are highly correlated. However, comparing current data with previous data values for finding redundant data does not show the greatest impacts of data transmission reductions, because of the instant deletion of previous data records after a successful transmission to the CH. While considering this, we proposed an attribute window concept w_M in every sensor node. However, w_M keeps the record of every 10 detected values during the observed time interval r and cleared the values on the basis of FIFO approach after each successful data window transmission to the CH node.

Furthermore, we also implement the A-DAWF mechanism in CHs for finding both SDRs and TDRs, since CHs receive the data messages of $SDR(\tau)$ among sensor nodes as well as their own sensor readings of $TDR(t)$. In CH nodes, we assume that the window w_Q size is 50, which can also be varied based on the CHs computational constraints for monitoring nodes spatial data and their own readings. In this case, if sensor nodes have the same application monitoring tasks at different periods, then there will be several SDRs over nodes. Hence, for reducing SDRs among nodes, CH keeps the spatial data records of its corresponding sensor nodes during every round time interval of R. In order to find SDRs, the following expressions Eq.(3) and (4) must be satisfied in CH nodes, and the parameter considerations and their ranges of SDRs are used almost same as like sensor nodes of Eq.(1) and (2) with exception of the window sizes, but the CH A-DAWF threshold η value is fixed at 0.05 and may vary since CHs window constraints are different than the SNs window constraints. According to the Eq.(3), SR_{count} represents a spatial correlations count of RV in CH windows.

$$SR_{count} = \begin{cases} if \frac{|q_{j+1}-q_j|}{q_j} > \eta, & SR_{count}= 0 \\ otherwise, & SR_{count}= 1 \end{cases} \quad (3)$$

$$\mu_L = \frac{\sum_{j=1}^{Q} q_{j+LQ}}{Q} \quad (4)$$

B. A-DAWF Prediction-Based Mechanism

The iterative Algorithm 1 starts by initializing all the variables and their parameters. The core of this algorithm, consists of a data redundancy loop that predicts the node redundant or correlated data with the previous data values. If the current readings are differ from the previous readings, then only node adds/counts the sorted data into its corresponding window packet before fowarding it to the cluster-head, otherwise the data can be flushed itself.

The entire algorithm iteratively stored the received data of their corresponding nodes into their attribute window functions, which can comput and update the data packets before forwarding to the CH node. In this algorithm, A-DAWF has two time intervals; one as nodes received data time intervals t and another one as window forwarding time intervals r. Using RV function (Algorithm 1, line 4) A-DAWF mechanism examines the window stored data of nodes, whether they are redundant and correlated, if RV is greater than the threshold then it counts and forwards the non-correlated data towards the BS. If the window contains the redundant information then (Algorithm, line 8) A-DAWF utilizes its mean averaging function to send one appropriate data value rather than the all window redundant messages. Moreover this pre-filtration method helps to reduce their redundant data transmissions, which then also reduces the total communication burden on nodes and maximizes the overall nodes lifetime as well as the network lifetime.

Moreover, we have also been developing the further considerations and improvements of the propositions in COOJA/Contiki simulator for both various metric and comparative analysis with several well known protocols. Since the simulator uses its own developed software's and can be uploaded directly on any recommended real motes. For

Algorithm 1 nodes sensed phenomena ($p_i(t)$)

procedure : $Initialize(n, t, r)$
Parameters: $t \leftarrow node\ time\ intervals$, $TR_{count} \leftarrow \frac{|p_{i+1}-p_i|}{p_i} > \eta$
$r \leftarrow window\ round\ time\ interval\ of\ the\ sensor\ nodes$

if $p_i(t)$ **then**
Store into the Window buffer w_M
if $\frac{abs(p_{i+1}-p_i)}{p_i} > \eta$ **then**
Count Temporal correlations
if $TR_{count}\ mod\ 10 == 0$ **then**
Add mean μ_K averaging data
else
Count non-correlated Data
Send as single data packet
end if
end if
end if
end procedure

instance in our case we consider Tmote-sky for simulations and TelosB devices for network deployment.

IV. Experimental Study and Preliminary Results

In this section, we presented the preliminary results of cluster topology with the help of real Intel Data sets from [12]. According to this online repository, the data values are captured every 31 seconds, and the readings (temperature, relative humidity and light) are collected from the indoor network deployment of 54 sensor nodes at Intel-Berkeley Research Labs, between February 28 and April 5, 2004. The preliminary study of this network simulations are permformed in MATLAB. In this experimental study case, we considered the temperature datasets for preliminary results, and we only used one day and night measures of three sensor nodes. However, while developing our proposition, we notice huge TDRs and SDRs with A-DAWF compared to the datasets sampling rate system as shown in Figure 2 and Figure 5.

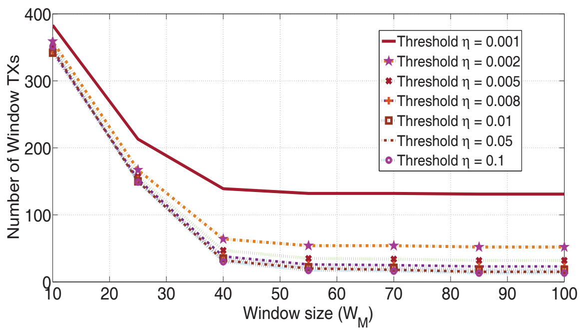

Fig. 2: Number of window transmissions (TX) through the various window sizes.

Figure 2 illustrates the total number of window data transmissions over the window w_M sizes at various error thresholds η as shown in the table I. Without A-DAWF, 1316 sensor readings would be sent by every node towards the network. In this scenario, we consider three cluster nodes which they do the TDRs filtration, where A-DAWF transmit 131 readings only by individual node to the CH out of 1316

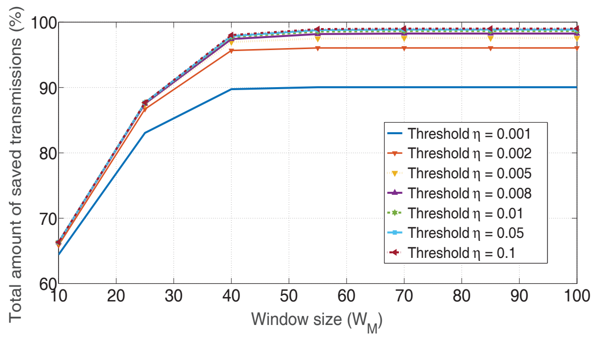

Fig. 3: Total number of saved transmissions (%) over the windows.

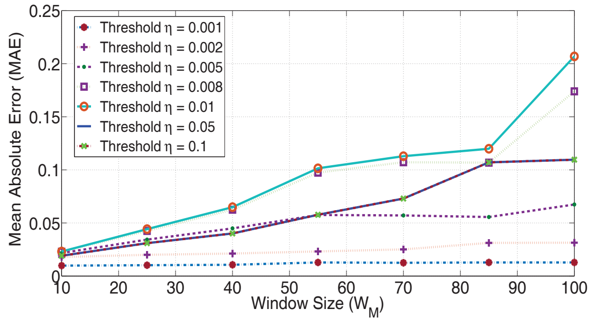

Fig. 4: Mean Absolute Error (MAE) cost over the window sizes.

readings compared to the datasets sampling rate. In order to calculate the transmission costs (TX_{cost}) in every node, we used a TX_{cost} metric of Eq.(5) to evaluate the performances of A-DAWF as shown in figure 3. However, we also conduct several simulation tests using different η and w_M parameters over the TX_{cost} and MAE metrics to explore the performances at both varying thresholds and window sizes as shown in figure 3 and figure 4. Figure 3 describes the total amount of saved transmissions over various window and threshold parameters, where a minimum window of 10 saves 66%, and a maximum window of 100 reduces more than 90% of the required transmissions . In figure 4, we present the mean absolute prediction error of sensor nodes at head node, which then shows the attained error over the various window sizes for minimum to maximum prediction error of the thresholds η are from 0.001 to 0.1. Following the minimum to maximum thresholds of MAE is less than one percent (<1%) in all cases.

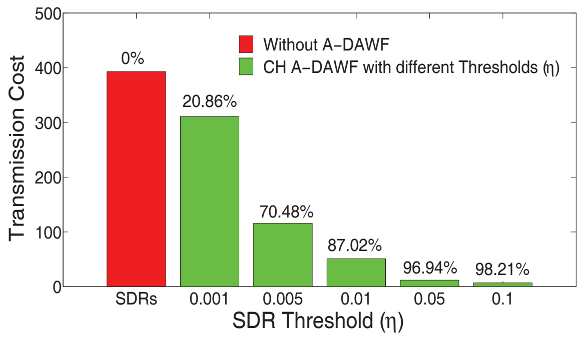

Fig. 5: CH A-DAWF among nodes spatial data with different Thresholds.

Figure 5 demonstrates the spatial correlations of data detection by a considered cluster nodes to the CH at different given time instances t from [12]. Which then figure 5 clearly shows that there are still large amount of spatial redundant data among the nodes, in which CH A-DAWF can still suppress a large amount of redundant data at a various thresholds and at a various window sizes of 10, 50 and 100. This experimental study help us to understand the differences and how often faulty readings generate the correlated values and their interactions between each other. Moreover, we also notice that there is no variation in TDRs and SDRs after η of 0.05 at a fixed window intervals, thus whatever the higher values are given then the A-DAWF deliverd data remain constant, which is obvious since RV only does the redundant data comparison between every two window data values, and process the non-redundant or mean averaging data accordingly. If the window size varies then the performances will obviously be varied, which are shown in the below tabulated readings.

TABLE I: A-DAWF performances at various thresholds and window sizes.

Thresh-old (η)	**window size (w)**	**Nodes Temporal data TXs (out of 1136)**
0.001	10	383
0.002	25	167
0.005	40	47
0.008	55	26
0.01	70	21
0.05	85	15
0.1	100	13

$$TX_{cost} = \left(1 - \frac{Num\ of\ window\ TXed\ messages}{Total\ num\ of\ periodic\ transmissions}\right) \times 100 \tag{5}$$

According to the TX_{cost} metric of Eq.(5), A-DAWF shows that it can suppress up to 90% of transmissions compared to the datasets sampling rate among the considered sensor nodes. On the other hand, CH node suppress 386 spatial redundancies or correlations of the nodes out of 393 data by a fixed optimal threshold η of 0.05 and at a window W_Q of 50, which means CH only sent 7 times among the nodes spatial data. Furthermore, we also considered to evaluate the message costs (M_{cost}) over nodes. According to the authors of [13] have experimented with two payload sizes of a node, such as 1 and 90 bytes of data carries towards the sink or BS to demonstrate that increasing the payload size (bytes) to some extent does not vary the energy consumption cost over nodes. In our considerations, A-DAWF sends its total uncorrelated window data as a single packet during its window time intervals.

The further in-depth research work of the algorithm, measurements and system analysis among the proposed network are work-in-progress, which can be presented in the future work proceedings.

V. Conclusion and Future Work

This paper is presented as work-in-progress implementations and the developed system models of spatio-temporal redundancy techniques for reducing TDRs in sensor nodes as well as both TDRs and SDRs in CH nodes. The preliminary results show that the proposed mechanism can suppress up to 90% of temporal redundant data among the considered sensor nodes by an optimal threshold of the window sizes as well as their spatial correlations are being suppressed effectively in a considered CH node compared to either periodic or a continuous transmission system. The ongoing measurements still need to be analyzed for measuring the message costs and overall network lifetime with the consideration of many CHs and their group nodes.

Further results and algorithm performances are being carried out with the consideration of our LE2P lab datasets from the recent network deployments. For comparative analysis, we consider to make a comparison between A-DAWF with our lab datsets and A-DAWF results from COOJA simulator. We are particularly interested to extend our propositions for monitoring different temporal physical properties at a time in nodes and their spatial correlations in CHs. Further experiments are also being examined in order to present comparative results with other well-known literature works to demonstrate the system performances especially in terms of quality of data, latency, message cost, and energy cost. Additionally, we also consider to design and develop A-DAWF mechanisms for multi-hop networks.

References

[1] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad hoc networks*, vol. 7, no. 3, pp. 537–568, 2009.

[2] S. Yoon and C. Shahabi, "Exploiting spatial correlation towards an energy efficient clustered aggregation technique (cag)[wireless sensor network applications]," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, vol. 5, pp. 3307–3313, IEEE, 2005.

[3] M. A. Sharaf, J. Beaver, A. Labrinidis, and P. K. Chrysanthis, "TiNA: a scheme for temporal coherency-aware in-network aggregation," in *Proceedings of the 3rd ACM international workshop on Data engineering for wireless and mobile access*, pp. 69–76, ACM, 2003.

[4] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on database systems (TODS)*, vol. 30, no. 1, pp. 122–173, 2005.

[5] N. Tsiftes, A. Dunkels, Z. He, and T. Voigt, "Enabling large-scale storage in sensor networks with the coffee file system," in *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, pp. 349–360, IEEE Computer Society, 2009.

[6] R. Tan, G. Xing, X. Liu, J. Yao, and Z. Yuan, "Adaptive calibration for fusion-based wireless sensor networks," in *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9, IEEE, 2010.

[7] A. De Paola, G. Lo Re, F. Milazzo, and M. Ortolani, "Predictive models for energy saving in wireless sensor networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*, pp. 1–6, IEEE, 2011.

[8] J. Wang, s. Tang, and B. Yin, "Data gathering in wireless sensor networks through intelligent compressive sensing," IEEE, Mar. 2012.

[9] Z. C. Taysi, M. A. Guvensan, and T. Melodia, "Tinyears: spying on house appliances with audio sensor nodes," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pp. 31–36, ACM, 2010.

[10] N. Tsiftes and A. Dunkels, "A database in every sensor," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pp. 316–332, ACM, 2011.

[11] S. Kandukuri, V. Ayadurai, J. Siden, and M. Prytz, *Power Control Mechanisms on WARP Boards.* Master Thesis Report, Mid Sweden University, Sundsvall, Sweden, Jan. 2013.

[12] S. Madden, "Intel Lab Data, http://db.csail.mit.edu/labdata/labdata.html.," 2004.

[13] C. Haas and J. Wilke, "Energy evaluations in wireless sensor networks: a reality check," in *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pp. 27–30, ACM, 2011.