# Distributed Adaptive Data Collection Protocol for Improving Lifetime in Periodic Sensor Networks

Ali Kadhum Idrees, Member, IAENG, and Ali Kadhum M. Al-Qurabat

***Abstract*—Periodic Sensor Networks (PSNs) represent one of the essential elements of emerging Cyber-Physical System (CPS) designs because of their using in multiple applications. PSNs are one of the major contributors of the big data in the future. One fundamental challenge in PSNs is to periodically collect the large volume of data in an energy efficient way and then transmit them to the sink so as to enhance the network lifetime. Since sensor batteries have a limited lifetime, therefore, adaptive sampling method to periodic data collection is required to support energy-efficient data gathering and fusion of CPS. In this research, we suggest a protocol, called Distributed Adaptive DAta Collection protocol (DADAC), which collects periodically sensor readings and prolong the lifetime of a Periodic Sensor Network (PSN). The lifetime of DADAC protocol is divided into cycles. Each cycle is composed of four stages. First, data collection. Second, dimensionality reduction using Piecewise Aggregate Approximation (PAA) technique to reduce the amount of data collected by each sensor. Third, Frequency Reduction using SAX (Symbolic Aggregate approXimation) approach in order to remove the redundant data before send them to sink. Fourth, sampling rate adaptation based PAA similarity to acclimate its rate of sampling according to the dynamic modification of the monitored environment. DADAC allows each sensor to remove the redundant collected data and adapts its sampling rate in accordance with the monitored environment conditions.**

**We conduct extensive simulation experiments on real sensor data by applying OMNeT++ network simulator to explain the effectiveness of the DADAC protocol in comparison with two other existing methods.**

***Index Terms*—Periodic Sensor Networks, Data Collection, Adaptive Sampling Rate, PAA similarity & SAX, Network Lifetime.**

## I. Introduction

In the last years, the Cyber-Physical System (CPS) has appeared as an important trend to enhance the interactions between virtual and physical worlds[1]. CPS combines between physical devices such as sensors with the cyber elements (i.e., informational) to form an intelligent system deals with dynamic changes of the physical environment. In many proposed studies, Wireless Sensor Network (WSN) is considered as a primary component of cyber physical systems. CPS is a combination between both cyber resources and WSN. WSN represents an important factor in people's life, because of their widespread use in many applications such as agricultural, healthcare, transportation, environment, industry, and military [2], [3], [4], [5]. WSN is composed of a large number of low-cost tiny sensors that deployed for monitoring physical phenomena of a specific region of interest such as temperature, sound, vibration, pressure, or motion [3]. Sensor node can sense, process, and communicate with limited capabilities in battery power, storage, computation, and bandwidth [4], [6], [7]. They collect the sensed data from the monitored environment, manipulate the data locally, and transmit them to the sink for further analysis [8].

One of the most critical constraints of the sensor node is the battery life. Due to the environment or cost restrictions, it is difficult or impossible to change or recharge the sensor batteries. Thus, the sensor nodes are deployed with high density in order to enhance the network lifetime. In sensor node, the radio unit represents the principal source of energy consumption. Therefore, it is important to remove redundant sensed data before reporting them to the sink to save the energy and improve the lifetime of sensor node [9], [10], [11]. It is necessary to take into consideration data capturing, communication, and routing problems in order to design energy-saving protocol for PSN. Data collection approaches determine the way of sensor's work in data collection and sending to the base station. Therefore, data collection represents the crucial function in PSNs [12]. The CPS gathers sensor readings from physical environment and joins them to different information sources for real-time analysis[13].

There are two models for data collection in WSNs: time-driven and event-driven [12]. This work considers time-driven data collection which is named Periodic Sensor Networks $(PSNs)$. In PSN, every sensor node transmits the sensed data of the monitored area to the sink periodically. Several PSNs applications use the periodic way to monitor certain conditions regularly such as pressure, humidity, temperature, etc. Two main challenges in PSN. First, PSN has to provide adequate lifetime in order to satisfy application's needs. Second, data management is more difficult due to the huge amount of collected data by this network. Many proposed works consider reducing the amount of data during the collection and communication without considering the accuracy of data. Data reduction aims to prolong the network lifetime and facilitate data analysis and decision making. In PSN, the change in the monitored environment can slow down or speed up. The energy consumption can be decreased when the sensor node modifies its sampling rate based on the dynamic modification of the monitored phenomena. Therefore, to prolong the network lifetime, adaptive sampling for periodic data collection is required for energy optimization and data reduction [14], [15].

This paper introduces the following contributions.

i) A protocol named DADAC is devised to collect the sensor data in an adaptive way such that the volume of data is reduced while PSN lifetime is enhanced. The principal idea of DADAC protocol is to utilize the similarity of collected data and adapts its sampling rate accordingly. DADAC works into cycles. Four stages in each cycle: data collection, dimensionality reduction,

Manuscript received 19 Jun 2017 ; revised 22 Jul 2017. This work was supported in part by the University of Babylon, Iraq.

Ali Kadhum Idrees is with the Department of Computer Science, University of Babylon, Iraq. e-mail: ali.idrees@uobabylon.edu.iq.

Ali Kadhum M. Al-Qurabat is with the Department of Software, University of Babylon, Iraq. e-mail: alik.m.alqurabat@uobabylon.edu.iq.

frequency reduction using SAX technique, and sampling rate adaptation using PAA similarity. The sensor node provides a new sampling rate after each cycle based on the similarity between the periods of one cycle.

ii) A new adaptive sampling rate algorithm based PAA similarity is suggested. In each cycle, the speed of readings capturing inside the sensor node depends mainly on the previously calculated sampling rate adaptively. DADAC protocol uses SAX approach to eliminate the redundancy in the collected measures before sending them the base station so as to save energy and improve the lifetime.

iii) The simulation results are accomplished by OMNeT++ network simulator to illustrate the effectiveness of the DADAC protocol. The DADAC protocol has been compared to two algorithms in the related works: PFF algorithm that proposed by Bahi et al.(2014) [16] and Harb et al. algorithm (2016) that introduced in [17].

The rest of this paper is organized as follows. Next section exhibits literature review. Section 3 explains the description of DADAC protocol. Protocol evaluation is shown in Section 4. Finally, we present the conclusion and future works in Section 5.

## II. Literature review

Adaptive collection approaches are considered as a good candidate to save energy and extend the network lifetime of PSNs. The major objective of an adaptive collection technique is to make the sensor node be able to change its sampling rate dynamically in accordance with the monitored environment conditions. This can reduce the repetitive gathered data, consume less energy, and decrease the processing load at the base station [14]. Adaptive collection avoids capturing the redundant samples by exploiting the correlation (temporal[18], [19], spatial[20], [21], or spatio-temporal[22], [23], [24]) between sensed data. In order to save the energy of a WSN, several methods are proposed in [25], [26], [27], [28], [29] to improve the lifetime of the network.

This section reviews some related literature concerning the adaptive data collection in WSNs. The works proposed in [20], [21] consider adaptive sampling schemes based spatial correlation among the physical sensed data. In [20], the sampling rate is adapted by the base station. Initially, the base station activates a set of sensors to get the sensed data of monitored environment. The correlation percentage is computed for the received sensed data to increase or decrease the activated sensors.

Some other approaches study temporal correlation among sensed data [18], [19]. Chatterja and Havinga [18] present a sampling algorithm based temporal correlation among sensed data. In this algorithm, the sampling rate is modified depending on the stability of the monitored environment. The sampling rate increases when the environment conditions are unstable, otherwise the rate decreases.

Spatio-temporal correlation is used by some adaptive sampling techniques such as in [22], [23], [24]. For instance, Masoum et al. [24] introduce an energy-saving mechanism for data collection. Their scheme exploits spatio-temporal correlation among sensors and their sensed data to determine the candidate sensors which are responsible for sampling and transmission. The selected sensors are adaptively changed.

Some researchers used prediction as a way to adjust the sampling rate of sensor nodes [30], [31], [32], [33], [34]. An energy saving information gathering scheme is proposed by Liu et al. [30] to predict the sampling rate inside sensor using ARIMA model. In [31], the authors presented an algorithm for adaptive sampling using Box-Jenkins approach to estimate the future sensor readings, depending on the existing readings. Alippi et al. [34] introduced a power aware adaptive sampling method for snow monitoring. Their algorithm provides online estimation based on fast Fourier transform.

In recent years, several adaptive sampling approaches in PSNs have been studied [35], [36], [37], [16], [14], [17]. Laiymani and Makhoul [35] proposed a scheme for adaptive sampling using ANOVA model and Fisher test in PSNs. This algorithm works at the sensor node to adapt its sampling rate. The authors in [16] proposed method to remove the repetition of collected data in PSN called Prefix Frequency Filtering (PFF). Makhoul et al. [14] suggested adaptive data gathering approach for PSN. They combine between ANOVA model and remaining energy to permit every sensor node to modify its sampling rate in accordance with environment dynamics. Srbinovski et al. [36] proposed a power saving data collection algorithm for power scavenging in WSNs. Their approach takes the energy harvesting from the monitored sensing area and modifies its sampling rate based on the remaining energy and observed environment. An adaptive sampling algorithm based on endocrine regulation mechanism $(EASA)$ in WSN is presented [37]. The EASA algorithm uses hormone information to control the nodes in working state or resting state and adjusts collection frequency dynamically. Harb and Makhoul [17] proposed adaptive data collection approach based set similarity among sensor readings. Their technique allows each sensor node to identify, first, the similarity between data collected among successive periods using set similarity function, then to adjust its sampling rate to the newly calculated score of similarity. The sensor node reduces the amount of redundant collected readings and extends the network lifetime.

This paper suggested a Distributed Adaptive DAta Collection (DADAC) protocol for PSNs. The major goal of DADAC is to remove redundant sensor readings, save energy, and improve the network lifetime. DADAC performs four main phases. First, data collection according to adaptive sampling rate. Second, PAA approach is applied to avoid the redundancy in the collected data. Third, DADAC allows to each sensor node to adapt its sampling rate for each cycle (cycle = 2 periods) based on the PAA similarity. Fourth, SAX technique is used to remove the repetition in the collected data and then transmits them to the sink. DADAC is simulated on the OMNeT++ network simulator using real data of sensor nodes. The comparison results show that our protocol can provide a better performance and prolong the network lifetime.

## III. Description of the DADAC protocol

DADAC protocol is given in more details in this section. The main objective of this protocol is to enable each sensor to modify its sampling rate adaptively in accordance with the dynamic changing of the monitored environment. Consequently, this reduces the amount of redundant gathered data and minimizes energy consumption (extend the PSN lifetime) whereas the quality of collected data is maintained sufficiently to allow significant analysis. Figure 1 illustrates the flowchart of the proposed DADAC protocol. This section describes in detail DADAC protocol stages and algorithms correlated with each stage. Table 1 explains some parameters used in this paper.

TABLE I
SOME PARAMETERS USED IN THIS PAPER

| | |
|---|---|
| $SMP_R$ | Sampling rate |
| $MIN_{SMP}$ | Application criticality |
| $S$ | Temperature readings series $S = s_1, ...., s_n$ |
| $S^p$ | PAA of S, $S^p$= $c_1^p, ...., c_w^p$ |
| $S^x$ | Symbolic representation of $S^p$, $S^x = c_1^x, ..., c_w^x$ |
| $w$ | PAA segments number to represent S |
| $a$ | Number of alphabet (for instance, if the alphabet = (w,x,y,z), $a$ = 4) |
| $\beta$ | Breakpoints, $\beta = \beta_1, ...., \beta_{a-1}$ |
| $n$ | Sensor id |
| $n_e$ | Remaining energy of sensor n |

## A. Data collection

Each sensor node senses the data reading periodically. These sensed time-arranged data readings set forms time series. Therefore, DADAC protocol treats the sensor readings as a time series. In our work, we named this time series as a temperature readings series.

DADAC protocol is periodic and works into cycles. The cycle includes two periods (j=2). The period is partitioned into time slots. Therefore, in this stage, the sensor node $n$ catches one temperature reading $s_i$ each time slot. At the end of each period, the temperature readings series of sensor $n$ is formed such that $S_n$ =$\{s_1, s_2, ..., s_{\rho-1}, s_\rho\}$, where $\rho$ is the total number of temperature readings captured during the period. The sensor node collect the temperature readings at $SMP_R$ speed. The $SMP_R$ is initiated to $\rho$ temperature readings per period. The redundant temperature readings captured by the sensor node increase in two states: short time slot and slowly variation of a monitored area of interest.

## B. Dimensionality reduction

Often, temperature readings series is too big to be analyzed and thus data approximation is necessary. The aim of data approximation is to minimize the volume of sensor readings while retaining the signal fundamental shape and characteristics. Since time series representation has a great impact on the simplicity and effectiveness of data readings mining; consequently, it is required to choose the suitable technique to represent the sensed readings series [38]. Several representation methods are found in the literature such as Discrete Fourier Transform (DFT), the Discrete Wavelet Transform (DWT), and Singular Value Decomposition (SVD)[39]. DADAC protocol uses a simple and efficient representation technique called PAA [40], [38].

In this stage, DADAC protocol converts the temperature readings series from its original form to the PAA representation in order to decrease the dimensionality of series. PAA divides this series into equal portions. After that, it calculates the mean for each portion. Data sorting is an integral part of data analysis [41]. It improves the search and merges the sequences efficiently. Therefore, to further improve the efficiency of PAA, the sensed temperature readings are sorted

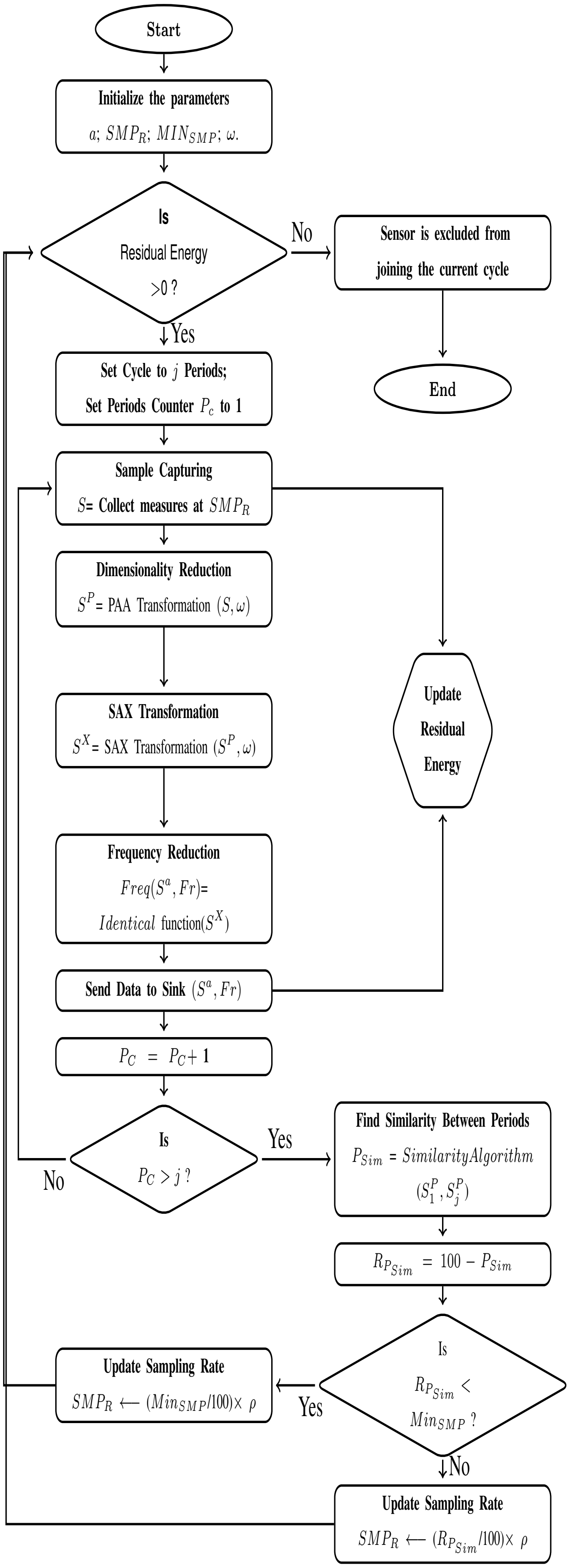



Fig. 1. Flowchart of the proposed DADAC protocol.

in descending order to group the similar ( or close similar) readings to each other. The sorted temperature readings series $S = \{s_1, ...., s_\rho\}$, is normalized using z-normalization before transforming it into PAA representation. This procedure guarantees that input temperature readings series converted to output series whose mean is nearly 0 whilst the standard deviation is in a range near to 1. This normalization is necessary to permit our algorithm to concentrate on the structural similarities/dissimilarities instead of the amplitude [39], [42]. This normalization is calculated as follow

$$\mu = \frac{\sum_{i=1}^{\rho}(s_i)}{\rho} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{\rho}(s_i - \mu)^2}{\rho - 1}} \tag{2}$$

$$s_i^{'} = \frac{s_i - \mu}{\sigma} \tag{3}$$

Where $\sigma$, $\mu$, and $s_i^{'}$ are the mean, standard deviation, and normalized temperature reading.

At each period, PAA represents the collected temperature readings series $S$ of length $\rho$ in a $w$-dimensional space (typically $w \ll \rho$) set such that $S^p$=$s_1^p$,... , $s_w^p$. Each $i^{th}$ value in $S^p$ is computed using the following formula

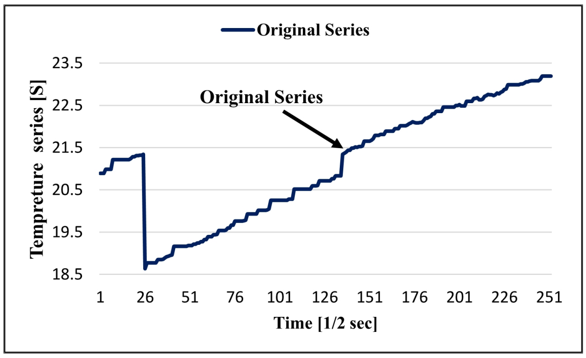
$$s_i^p = \frac{w}{\rho} \sum_{j=\frac{\rho}{w}(i-1)+1}^{\frac{\rho}{w}i} s_j^p \tag{4}$$

In order to decrease the dimensionality of temperature readings series from $\rho$ to $w$, the sensed temperature readings series is partitioned into segments with equal sizes. After that, PAA computes the mean for the sensed data readings within each segment so as to produce a sensed data-reduced representation set. The time complexity of computing the mean in Eq. 4 is $O(\rho)$. The transformation process of original temperature readings series to PAA representation is presented in Algorithm 1.

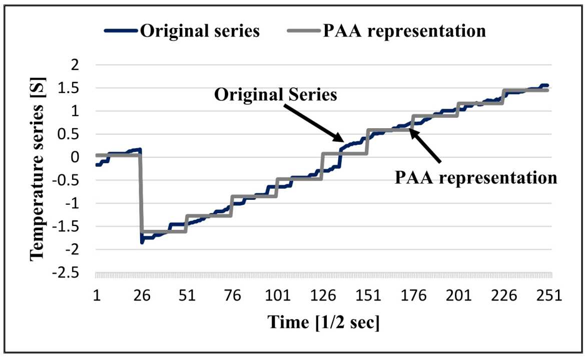
**Algorithm 1.** PAA Dimensionality Reduction

**Require:** $S$ ($\rho$-dimensional temperature readings series), $w$ (PAA segments number)
**Ensure:** $S^p$ (set of PAA Coefficients)
1 : $S \longleftarrow$ Sorting($S$) in descending order
2 : **For** $i \longleftarrow 1$ to $\rho$ **do**
3 : $\quad s_i^p \longleftarrow \frac{s_i - \mu}{\sigma}$
4 : **end for**
5 : **For** $k \longleftarrow 1$ to $w$
6 : $\quad {S^p}_k \longleftarrow \frac{w}{\rho} \sum_{j=\frac{\rho}{w}(k-1)+1}^{\frac{\rho}{w}k} s_j^p$
7 : **end for**
8 : return $S^p$

Figure 2 depicts PAA transformation process of the original temperature readings series to PAA representation. As shown in this figure, The temperature readings of 250 length is decreased to 10 dimensions.




(a)




(b)

Fig. 2. PAA transformation process (a) original temperature readings series and (b) PAA representation.

## C. Frequency reduction

Since radio component represents the most energy-consuming element in the sensor node, therefore, it is necessary to report as less as possible of sensed readings to the base station so as to extend the PSN lifetime while maintaining the accuracy of transmitted sensed readings. The main objective of this stage is to minimize the number of temperature readings which are gathered by each sensor node and maintain the frequency for each reading so as to not influence on the readings analysis in the base station. In this stage, DADAC protocol uses SAX representation to eliminate the redundant sensed temperature readings before sending them to the base station. This can save energy and improve the PSN lifetime. SAX method is one of the first symbolic representation that reduces dimensionality/numerosity of time series [39]. It converts the data series into a set of symbols. Each symbol takes its value from a finite alphabet [43].

The PAA representation $S^p$ of the original temperature readings series $S$ is transformed into SAX representation $S^x$ using the following steps:

i) Partition the temperature readings series into $w$ portions.
ii) The mean for each portion readings is computed.
iii) The mean values are quantized into symbols selected from an alphabet of size $N$.

The first two steps are PAA representation. In step 3, The quantization uses $(N-1)$ breakpoints which partition the region under the Gaussian distribution into $a$ equal proportional regions. Breakpoints can be defined as a sorted values list $B = \beta_1, ...., \beta_{a-1}$. The region under a $N(0,1)$ Gaussian curve from $\beta_i$ to $\beta_{i+1}$ = $1/a$, where $\beta_0$ and $\beta_a$ refer to $-\infty$ and $\infty$, respectively. The breakpoints are located by search them in a statistical table. For instance, Table 2 shows A lookup table of the breakpoints for $a$ with values range from 3 to 10 [39].

TABLE II
A LOOKUP TABLE OF THE BREAKPOINTS FOR $a$.

| | $a$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $\beta_i$ | $\beta_1$ | **-0.43** | **-0.67** | **-0.84** | **-0.97** | **-1.07** | **-1.15** | **-1.22** | **-1.28** |
| | $\beta_2$ | **0.43** | **0** | **-0.25** | **-0.43** | **-0.57** | **-0.67** | **-0.76** | **-0.84** |
| | $\beta_3$ | | **0.67** | **0.25** | **0** | **-0.18** | **-0.32** | **-0.43** | **-0.52** |
| | $\beta_4$ | | | **0.84** | **0.43** | **0.18** | **0** | **-0.14** | **-0.25** |
| | $\beta_5$ | | | | **0.97** | **0.57** | **0.32** | **0.14** | **0** |
| | $\beta_6$ | | | | | **1.07** | **0.67** | **0.43** | **0.25** |
| | $\beta_7$ | | | | | | **1.15** | **0.76** | **0.52** |
| | $\beta_8$ | | | | | | | **1.22** | **0.84** |
| | $\beta_9$ | | | | | | | | **1.28** |

When the breakpoints have been determined, the PAA coefficients can be quantized as follow. Every PAA value less than the smallest breakpoint will be converted to "a" symbol, whilst the PAA values that are equal to or larger than the smallest breakpoint and less than the second smallest breakpoint are converted into "b" symbol, etc. Figure 3 presents mapping PAA values into SAX symbols using breakpoints, where $\rho$ =250, $w$ = 10, and $a$ = 5. In this figure, the temperature readings series is mapped to the word "caabbcdeee".

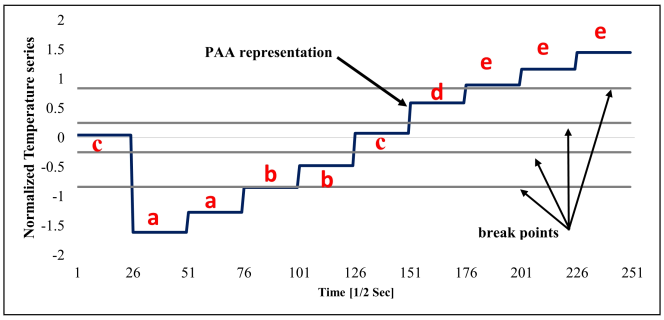



Fig. 3. Mapping PAA values into SAX symbols using breakpoints.

In Figure 3, SAX representation provides five symbols: a, b, c, d, and e. The symbols can be merged to introduce a sequence called word. It can be defined as follow. Let $alpha_i$ indicates the $i^{th}$ value of the alphabet ( i.e., $alpha_1$ = a and $alpha_2$= b). Consequently, the transformation from a PAA representation $S^p$ to a word $S^x$ is computed as follows

$$S^x{}_i = alpha_j, \qquad iif \quad \beta_{j-1} \le s^p{}_i < \beta_j. \tag{5}$$

After converting the PAA values into SAX symbols, the resulted SAX symbols sequence will include redundant symbols due to multiple consecutive segments are transformed to the same symbol. In this stage, DADAC protocol removes these redundant symbols in each period to prevent transmitting the same symbols to the base station. Therefore, we will define a function that allows each sensor node to find the similarity among the symbols of *word* $S^x = s_1^x, ..., s_w^x$ to eliminate this redundancy.

The identical function identifies the similarity between two symbols $s_i^x$ and $s_j^x$ and can be defined as follow

$$Identical(s_i^x, s_j^x) = \begin{cases} 1 & if \quad s_i^x = s_j^x \\ 0 & otherwise. \end{cases}$$

The sensor node $n$ will search for the same symbols in the $word$ of the period $j$. If the same symbols are found, the sensor will record only the first occurrence of the symbol, and removes the others while increasing the frequency of the symbol by one every time occurs in the word. Otherwise, the sensor will add this symbol to the set and initiated its frequency to 1. Frequency of the Symbol $fr(s_i^x)$ is defined as the number of the occurrence of the same symbol in the same set. This reduced set of symbols and its frequencies will be transmitted to the base station after converting each symbol to its equivalent mean of PAA segment. The process of SAX frequency reduction is illustrated in algorithm 2.

**Algorithm 2.** SAX Frequency Reduction

**Require:** $S^p$ ($w$-dimensional PAA coefficients), $a$ (Alphabet Length), $\alpha$ (Alphabetic)
**Ensure:** $R^s$ (Reduced set of readings), $F^{R^s}$ (Frequencies of readings in $R^s$)
**For** $i \longleftarrow 1$ to $w$ **do**
  **For** $j \longleftarrow 1$ to $a$ **do**
    **if** $\beta_j \le s_i^p < \beta_{j+1}$ **then**
      $S^x{}_i \longleftarrow \alpha_j$
    **end if**
  **end for**
**end for**
$z \longleftarrow 0$
**For** $m \longleftarrow 1$ to $a$ **do**
  $z \longleftarrow z + 1$
  **For** $n \longleftarrow 1$ to $w$ **do**
    **if** $Identical(S^x{}_n, \alpha_m) = 1$ **then**
      $F_z^{R^s} \longleftarrow F_z^{R^s} + 1$
    **end if**
  **end for**
  **if** $(F_z^{R^s} > 1)$ **then**
    $R_z^s \longleftarrow \alpha_m$
  **end if**
**end for**
**For** $i \longleftarrow 1$ to $z$ **do**
  $d \longleftarrow Ascii(R_i^s) - 97$
  $R_i^s \longleftarrow \beta_d \times \sigma + \mu$
**end for**
return $R^s$, $F^{R^s}$

*D. Adaptive sampling rate*

In this stage, DADAC protocol modifies its sampling rate based on the percentage of similarity between temperature readings of different periods in the cycle. The main purpose of this stage is to calculate the similarity among periods after each finished cycle to acclimate the rate of sampling according to the new similarity rate. DADAC protocol adapts its rate of sampling at the end of each cycle. Therefore, it uses the PAA distance measure to find the amount of similarity between periods of each cycle.

*1) Similarity measure:* This section describes some measures used to give the distance between two temperature readings series. *Euclidean* distance represents one of the most famous distance measures. *Euclidean* distance between two temperature readings series $Q$ and $C$ of the same length $\rho$ is given by the following formula

$$D_E(Q, C) = \sqrt{\sum_{i=1}^{\rho} (q_i - c_i)^2}. \tag{6}$$

Let $Q$ and $C$ represent two collected temperature readings series. The PAA representations for the two series are $Q^p$ and $C^p$. The lower bounding approximation of the *Euclidean* distance between the two temperature readings series $Q^p$ and $C^p$ can be obtained using the following formula

$$D_{PAA}(Q^p, C^p) = \sqrt{\frac{\rho}{w}} \sqrt{\sum_{i=1}^{w} (q_i^p - c_i^p)^2}. \tag{7}$$

An examples for the visual representation of distance measures presented in Eq. 6 and Eq. 7 can be shown in[39].

DADAC protocol uses *Similar* function to identify the *Similarity* between two transformed PAA temperature readings series $Q^p$ and $C^p$. The *Similar* function refers to the similarity between two PAA temperature readings series using the following formula

$$SIM(Q^p, C^p) = \frac{1}{1 + D_{PAA}(Q^p, C^p)}. \tag{8}$$

After that, in order to measure the similarity percentage ($P_{Sim}$), it is defined as follow

$$P_{Sim} = SIM(Q^p, C^p) \times 100. \tag{9}$$

Algorithm 3 gives the similarity percentage ($P_{Sim}$) calculation between two PAA temperature readings series $Q^p$ and $C^p$.

**Algorithm 3.** Similarity Algorithm

**Require:** $Q^p$, $C^p$, $\rho$ (dimension before reduction)
**Ensure:** $P_{Sim}$
$Sum \longleftarrow 0$
**For** $i \longleftarrow 1$ to $w$ **do**
  $Sum \longleftarrow Sum +$ $({Q^p}_i - {C^p}_i)^2$
**end for**
**if** $(Sum = 0)$**then**
  $D_{PAA} \longleftarrow 0$
**else**
  $D_{PAA} \longleftarrow \sqrt{\frac{\rho}{w}} \sqrt{Sum}$
**end if**
$Sim \longleftarrow \frac{1}{1+D_{PAA}}$
$P_{Sim} \longleftarrow Sim \times 100$
return $P_{Sim}$

*2) Verification the similarity of periods:* In DADAC protocol, the sampling period refers to the time duration during which the sensor capture sensed temperature readings from the surrounding environment. The speed of change of environmental conditions and what fundamental features should be periodically gathered in temperature readings collection model can influence on the sampling period. In DADAC protocol, every node able to adapt its rate of sampling according to the amount of similarity among temperature readings series collected during different periods. The aim of computing the similarity between the temperature readings series every cycle is to adapt the rate of sampling based on the new calculated similarity. Therefore, the PAA similarity coefficient is employed to discover the similarity percentage, $P_{Sim}$ among several periods per cycle. On one hand, if $P_{Sim}$ is high, it means the monitored condition is changed at a slow speed. Therefore, the sensor node will decrease its rate of sampling to the minimum value to prevent collecting redundant readings. On the other hand, if $P_{Sim}$ is low, the sensor node will collect temperature readings at approximately maximum sampling rate so as to prevent losing significant readings. Therefore, to acclimate the rate of sampling of sensor node in accordance with the computed similarity among periods, the reverse of similarity percentage for PAA similarity coefficient $R_{P_{Sim}}$ is computed as follows

$$R_{P_{Sim}} = 100 - P_{Sim}. \tag{10}$$

Consequently, the computed $R_{P_{Sim}}$ will be used to acclimate the rate of sampling of the sensor in the new periods. When there is a high degree of similarity among periods (i.e., $P_{Sim}$ is high), the sensor node balances its rate of sampling to the minimum value ($R_{P_{Sim}}$ is low). Otherwise, it balances its rate of sampling to the maximum value. As aforementioned, the process of adapting the sampling rate in the sensor node depends on the $R_{P_{Sim}}$, thus the application criticality will be taken into consideration in this process.

*3) Application criticality:* The PSN can be used for monitoring disasters by using various kinds of sensor devices, e.g., for temperature, displacement, pressure, and concentration of chemicals, or noise detection. The influence of disasters on people and on the environment is not the same. Therefore, the sensor can modifies its rate of sampling in a different manner for each monitored disasters. In order that, if the risk level of the disaster is high then the sensor node must collect sensed readings more than if the risk level of the disaster is low. This can provide collected readings with high quality to make both of the analysis easier, and the monitored disaster is better to understand. There is an inversely proportional relation between $P_{Sim}$ and $R_{P_{Sim}}$. Therefore, when the similarity among periods is high, the $R_{P_{Sim}}$ will force the sensor node to make its sampling rate as minimum as possible.

In general, when the sensor node has the ability to alter its rate of sampling depending on the application's needs in PSNs, this will save its energy. In DADAC protocol, the criticality of application is expressed as a minimum amount of sampling rate in a period for a sensor node, $MIN_{SMP}$. $MIN_{SMP}$ takes values in the range 0 to 100 which represent the criticality level either low or high respectively. The sensor node adapts the new sampling rate to the $MIN_{SMP}$ (not to the $R_{P_{Sim}}$) when the recently calculated sampling rate is less than $MIN_{SMP}$. Depending on the requirements of the application and before the deployment, all the sensor nodes initialize their $MIN_{SMP}$. It is also possible to change $MIN_{SMP}$ dynamically during the lifetime of the network for the whole sensors or for just a given subgroup of sensors if there are some types of management and control schemes are available.

Algorithm 4 illustrates an adaptive sampling rate approach. The main purpose of this algorithm is to give every sensor device the ability to modify its rate of sampling to conserve

**Algorithm 4.** Adaptive Sampling Rate Algorithm

**Require:** $j$ (One cycle = $j$ periods), $\rho$, $MIN_{SMP}$, $a$: alphabet
**Ensure:** $SMP_R$ (new sampling rate)
1 : $SMP_R \longleftarrow \rho$
2 : **while** $n_e > 0$ **do**
3 : **for** $i \longleftarrow 1$ to $j$ **do**
4 : Collect readings series ($S_i$) at $SMP_R$ speed
5 : $S_i^p \longleftarrow$ Algorithm1 ($S_i, w$)
6 : $SendToSink(R^s, F^{R^s}) \longleftarrow$ Algorithm2 ($S_i^p, a, \alpha$)
7 : **end for**
8 : **for** each cycle **do**
9 : $P_{Sim} \longleftarrow$ Algorithm3 ($S_1^p, S_j^p, \rho$)
10 : $R_{P_{Sim}} \longleftarrow 100 - P_{Sim}$
11 : **if** $R_{P_{Sim}} < MIN_{SMP}$ **then**
12 : $SMP_R \longleftarrow (MIN_{SMP}/100) \times \rho$
13 : **else**
14 : $SMP_R \longleftarrow (R_{P_{Sim}}/100) \times \rho$
15 : **end if**
16 : **end for**
17 : **end while**

its power and to decrease the volume of collected data. Algorithm 4 works into cycles and each cycle consists of $j$ periods. In each period, the sensor captures $\rho$ temperature readings. The number of periods $j$ is fixed to 2 in algorithm 4. For each cycle, the sensor node looks for the similarity percentage among periods (line 9), then it computes $R_{P_{Sim}}$ (line 10). Therefore, the sensor node will decide to increase its sampling rate to computed $R_{P_{Sim}}$ when it is greater than $MIN_{SMP}$ which is determined by the application. Otherwise, it decreases its sampling rate to the $MIN_{SMP}$ (lines 11-15).

## IV. Protocol Evaluation

### A. Simulation framework

To study and evaluate DADAC protocol, extensive simulations are performed with discrete event simulator OMNeT++ [44]. DADAC protocol is distributed at each sensor node and it is based on the dataset of Intel Berkeley Research Lab [45]. PSN in this Lab includes 54 Mica2Dot sensors. The sensed data of the weather (such as temperature, humidity, and light) are periodically collected by these sensors once each 31 seconds. The base station is located at the center of the Lab. It receives sensed readings from each sensor node by a single hop. In our simulation, the sensor nodes use a log file contains about 2.3 million readings collected previously by Mica2Dot sensor nodes in the Lab. This article uses only one measure of sensor node measurements: $temperature$[1]. Some sensor nodes are not used in our simulation because its data may be missing or truncated. Therefore, the temperature readings of 47 sensor nodes are selected and stored. The results are the average of 47 sensor nodes. Table III gives the selected parameters settings.

In the experimental simulations, Some performance metrics are applied to assess the effectiveness of the DADAC protocol such as sampling rate adaptation, number of collected temperature readings by a sensor node, number of sent temperature readings, energy consumption, and lifetime.

DADAC protocol uses the same energy consumption model discussed in [17]. Energy consumed by the sensor

[1]the others are done by the same manner.

TABLE III
SIMULATION PARAMETERS FOR PSN INITIALIZATION

| Parameter | Value |
|---|---|
| PSN size | 47 nodes |
| $a$ | 5 and 10 symbols |
| $\rho$ | 20, 50, 100 and 200 readings |
| $MIN_{SMP}$ | 20, 40 and 60 |
| $w$ | $10\%, 20\%, 50\%, 75\%$ of the collected measures |
| $j$ | 2 |
| $E_{elec}$ | 50 nJ/bit |
| $\beta_{amp}$ | 100 pJ/bit/$m^2$ |

node depends only on the periodically collected and sent temperature readings to the base station. The cost of transmission is calculated for a $m - bits$ message and for a distance $d$ as follow

$$E_{TX}(m,d) = E_{elec} * m + \beta_{amp} * m * d^2. \qquad (11)$$

The energy consumption required for capturing $m - bits$ by the sensor node is calculated as follow

$$E_{CX}(m,d) = E_{TX}(m,d)/7. \qquad (12)$$

These experiment simulations consider the length of data reading $m$ equal to 64. In the case of transmission, 16 bits are added to $m - bits$ message which corresponds to the frequency of data reading $m$. Consequently, Energy consumption is defined as the total energy dissipated at each sensor node during the collection and transmission of data readings and formulated as follow

$$E_{Total} = E_{TX}(m,d) + E_{CX}(m,d). \qquad (13)$$

### B. Performance analysis

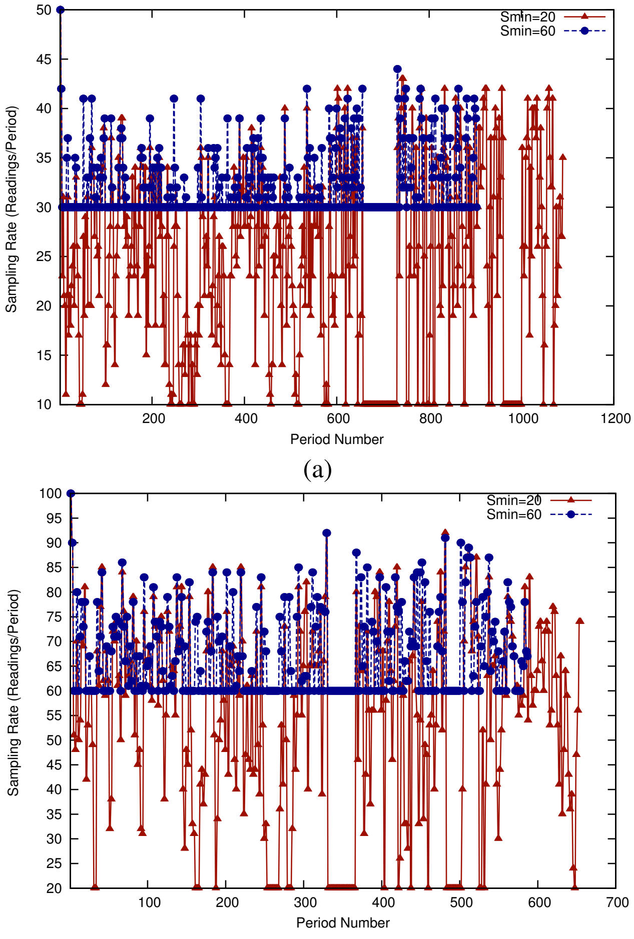
In this section, several experiments are achieved to show the performance of DADAC protocol. It is distributed at each sensor node in the PSN. Every node reads real temperature readings periodically and adapts its rate of sampling after each cycle based on the similarity percentage among collected sets of temperature readings.

*1) Sampling rate adaptation:* Figure 4 shows the adaptation of sampling rate and for two sizes of temperature readings (50 and 100 respectively). The results illustrate the ability of sensor device to modify its rate of sampling dynamically depending on the application criticality level. The risk level $MIN_{SMP}$ can be determined according to the type and requirement of application used to monitor the disaster.

In this experiment, $MIN_{SMP}$ uses two values: 20 for low risk level disaster and 60 for high risk level disaster. As shown in Figure 4, the adaptation of sampling rate is dynamic and after each cycle based on the application criticality level (i.e., $MIN_{SMP}$= 20 or 60). The results in Figure 4 (a) and (b) validate the good performance of our protocol.

*2) Number of collected readings:* Figure 5 shows the number of collected readings by the node at the end of simulation. DADAC protocol uses different values for the parameters $SMP_R$, $a$, $MIN_{SMP}$, and $w$.

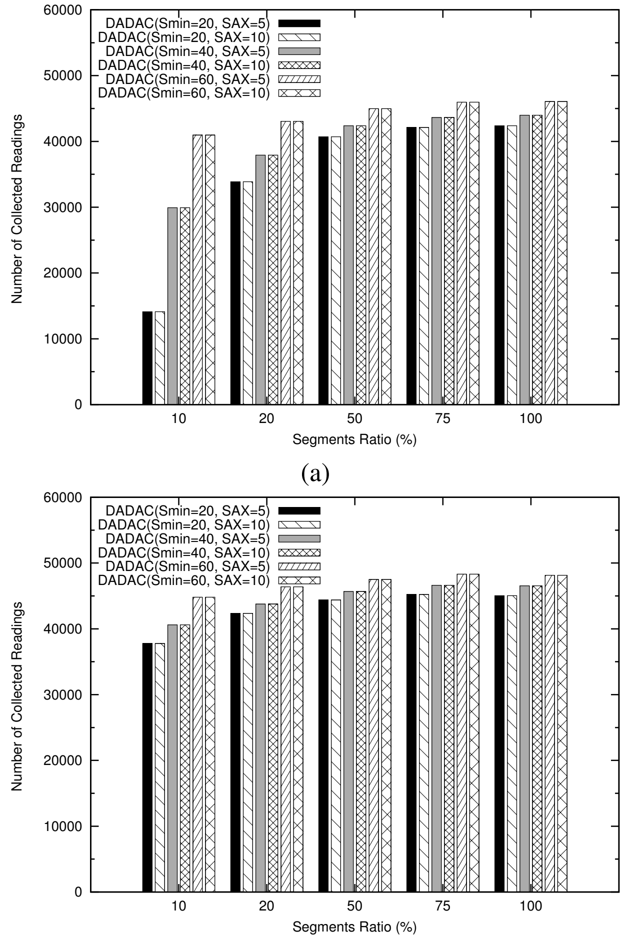
As shown in these results, the alphabet size $a$ does not affect the number of collected readings because of adaptation of sampling rate depends basically on the similarity among periods. DADAC protocol collects as large as possible of




(a)

(b)

Fig. 4. Sampling rate adaptation (a) $\rho$ = 50 and (b) $\rho$ = 100.




(a)

(b)

Fig. 5. Number of collected readings (a) $\rho$ = 50 and (b) $\rho$ = 100.

temperature reading as the $MIN_{SMP}$ increases. This can support application requirements when the risk level is high then DADAC protocol collect more readings. It can be seen that increasing the $\rho$ leads to increase the number of collected readings because of the decreasing in the similarity percentage between collected readings of successive periods. Concerning the accuracy, it can be seen that when the number of segments $w$ decreases then the number of collected readings decreases due to increasing the similarity among collected readings according to the defined identical function.

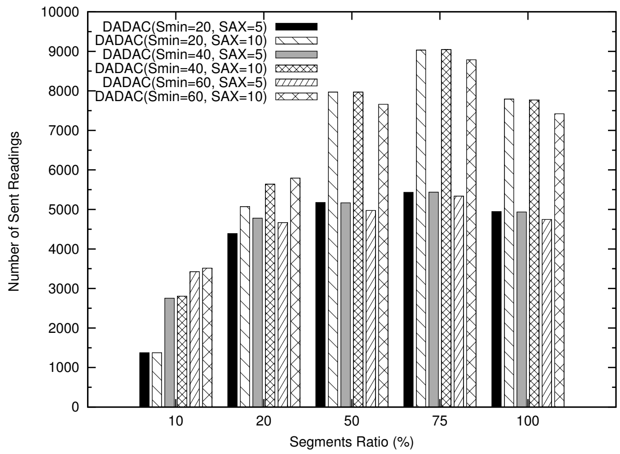
*3) Number of sent readings:* In this experiment, the number of sent readings by sensor node is evaluated. Another task carried out by DADAC protocol is to remove redundant collected readings before send them to the base station while maintaining the accuracy of collected readings. Figure 6 indicates the number of sent readings by the node at the end of the simulation.

Obviously, the number of sent readings increases with the number of alphabet sizes. This is due to the lack of similarity among collected readings. It can be seen that DADAC protocol send the larger amount of readings to the base station when the $MIN_{SMP}$ increases. This can support the application needs by sending a larger number of readings when the risk level of application is high. It is obvious that the increase in the $SMP_R$ leads to decrease the number of sent measures due to SAX transformation of collected readings into fixed number of symbols, each one associated with different frequency. For example, suppose $a$ =5 and $SMP_R$=50, the 50 collected readings will be represented by 5 symbols (a, b, c, d, and e). Each of these symbols has a different associated frequency (e.g., 5, 4, 15, 10, 16). If the $SMP_R$ increases to 100 for the same 5 symbols, it leads to represent the 100 collected readings by the same 5 symbols and with different associated frequency for each symbol. Therefore, DADAC protocol reduces the number of redundant data before send them to the base station to saves more energy and improve lifetime.

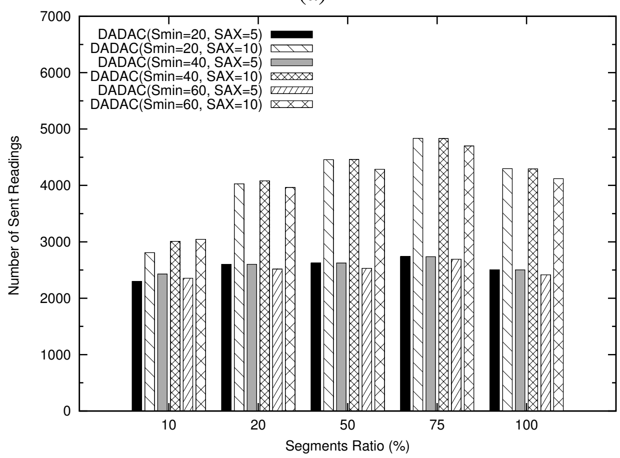
*4) Energy consumption:* In this experiment, the energy consumption of the sensor node using DADAC protocol is studied. Figure 7 illustrates energy consumption by a sensor node at the end of the simulation.

As shown in Figure 7, when the $a$ and $w$ increase, the number of sent readings increases (see Figure 6) thus energy consumption by the sensor node using DADAC protocol increases. DADAC protocol increases the sent readings when the risk level of the application is high. Therefore, the energy consumption by DADAC protocol increases when the $MIN_{SMP}$ increases. Furthermore, it is obvious that the increase in the $SMP_R$ leads to decrease the number of sent readings thus save the energy of sensor node.

*5) Data Accuracy:* This experiment shows another significant evaluation metric to assess the quality of DADAC protocol is the data accuracy of collected data. This metric refers to the rate of the data lost. However, the data accuracy is computed at the end of the experiment by subtracting the

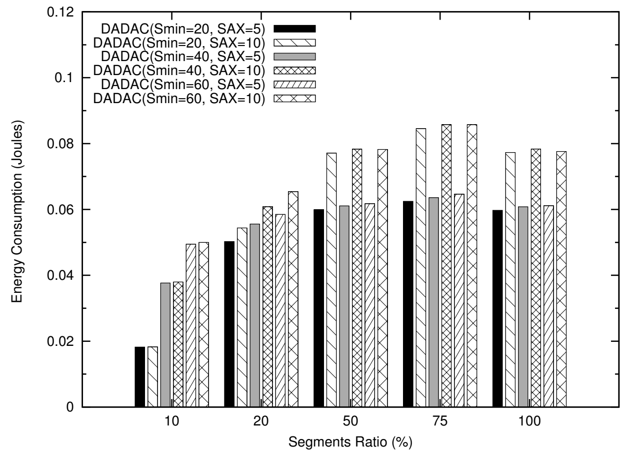



(a)

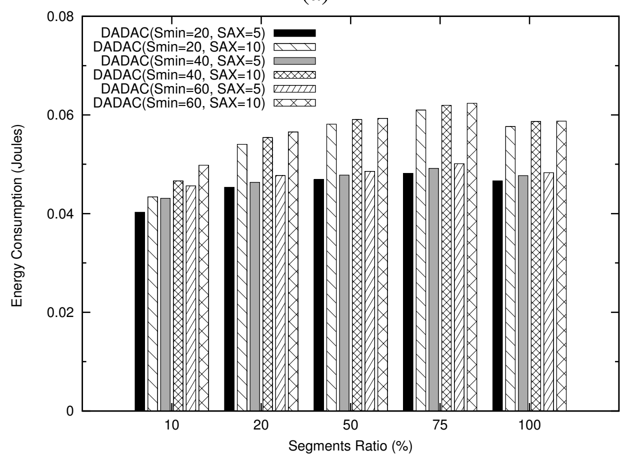



(b)

Fig. 6. Number of sent readings (a) $\rho$ = 50 and (b) $\rho$ = 100.




(a)




(b)

Fig. 7. Energy consumption by a sensor node (a) $\rho$ = 50 and (b) $\rho$ = 100.

lost data readings rate from the total number of data readings collected by the sensor without adaptive sampling. Tables IV and V shows the data accuracy of DADAC protocol for $\rho$ = 50 and $\rho$ = 100 respectively with $a$ = 5. The conducted results confirm that the DADAC protocol gives a suitable level of data accuracy. It gives at least 99.34 % of data accuracy. Hence, the decision at the base station will be not affected. Therefore, DADAC protocol can be considered as an energy-efficient method to adapt the rate of sampling of the sensor whilst keeping a high level of data accuracy of the gathered data.

TABLE IV
DATA ACCURACY OF COLLECTED DATA ($\rho$ = 50).

| Segments Ratio (%) | Data Accuracy (%) | | |
|---|---|---|---|
| | Smin=20 | Smin=40 | Smin=60 |
| 10 | 99.36 | 99.34 | 99.64 |
| 20 | 99.39 | 99.59 | 99.69 |
| 50 | 99.56 | 99.66 | 99.73 |
| 75 | 99.34 | 99.54 | 99.67 |

### C. Comparison results

Depending on the conducted results in the subsection IV-B, DADAC protocol, with $a$ = 5 and segments ratio = 10%, seem to give the best results to be compared with the best results of other two existing techniques. The first scheme is called PFF that proposed by Bahi et al.(2014) [16]. The second approach is called Harb et al. (2016) that introduced in [17].

TABLE V
DATA ACCURACY OF COLLECTED DATA ($\rho$ = 100)

| Segments Ratio (%) | Data Accuracy (%) | | |
|---|---|---|---|
| | Smin=20 | Smin=40 | Smin=60 |
| 10 | 99.49 | 99.58 | 99.64 |
| 20 | 99.56 | 99.61 | 99.66 |
| 50 | 99.54 | 99.57 | 99.62 |
| 75 | 99.57 | 99.60 | 99.63 |

*1) Number of collected readings:* Figure 8 illustrates the number of collected readings at the end of simulation by every sensor node using DADAC protocol compared with other two approaches. DADAC protocol decreases the number of collected readings by a sensor node from 18% to 76% compared to PFF. The PFF does not allow to the sensor node to adapt its sampling rate. Therefore, it always collects the same number of readings. Whereas, in comparison with Harb et al. approach which allows the sensor node to adapt its rate of sampling based on the similarity between the periods of one cycle, DADAC protocol decreases the collected readings from 5% to 29%.

The results illustrate that DADAC protocol has the ability to get rid of the redundant collected readings efficiently so as to decrease the overhead of transmitted readings to the base station thus improve the network lifetime. It can be seen that DADAC protocol increases the volume of collected

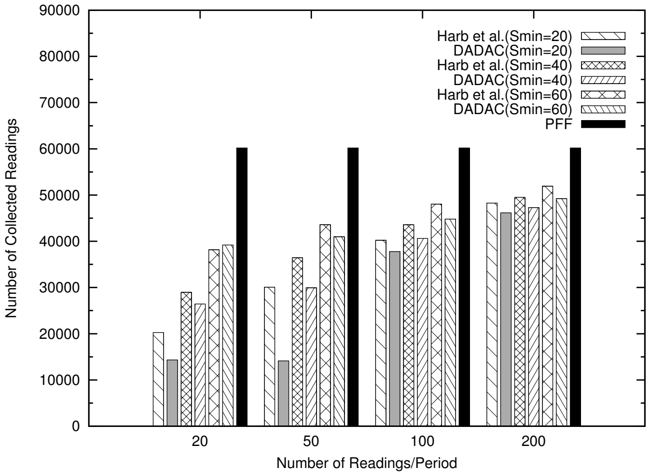



Fig. 8. Number of collected readings by a sensor node.

readings when the $MIN_{SMP}$ is increased. This increment in the collected readings is to meet the application requirements when the risk level is high.

*2) Number of sent readings:* When collecting the data readings at each period, DADAC protocol at the sensor node able to decrease the number of sent readings to the base station by using SAX method. Therefore, DADAC protocol finds the redundant symbols in the $word$ of each period and allocates for every symbol its frequency. Figure 9 demonstrates the number of sent readings by a sensor node to the base station at the end of simulation for DADAC protocol compared with the PFF and Harb et al. methods.

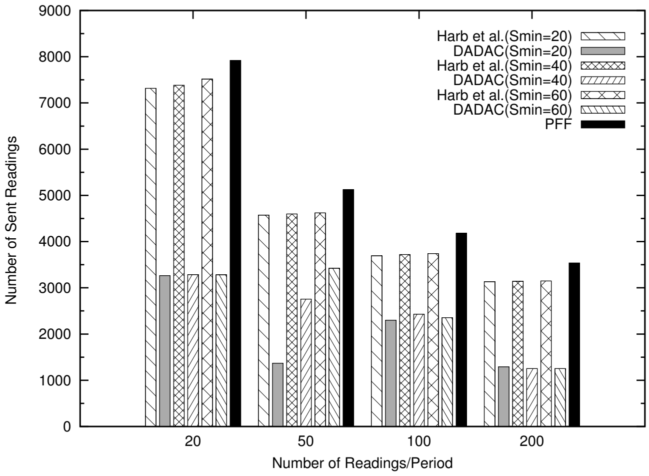



Fig. 9. Number of sent readings by a sensor node.

The results illustrate that DADAC protocol at each sensor node decreases from $56\%$ to $65\%$ of the number of sent readings to the base station as comparing to the PFF and from $55\%$ to $60\%$ as comparing to the Harb et al. methods respectively. Therefore, DADAC protocol removes the redundant collected readings successfully and the number of sent readings to the base station is reduced. We can also see that the volume of sent readings from the sensor node to the base station decreases when $\rho$ increases or $w$ decreases. This is due to the number of sent readings rely on the number of collected readings, segments ratio, the identical function, and the risk level of application.

*3) Energy consumption:* Figure 10 shows the energy consuming by DADAC protocol at the sensor node compared with PFF and Harb et al. approaches.

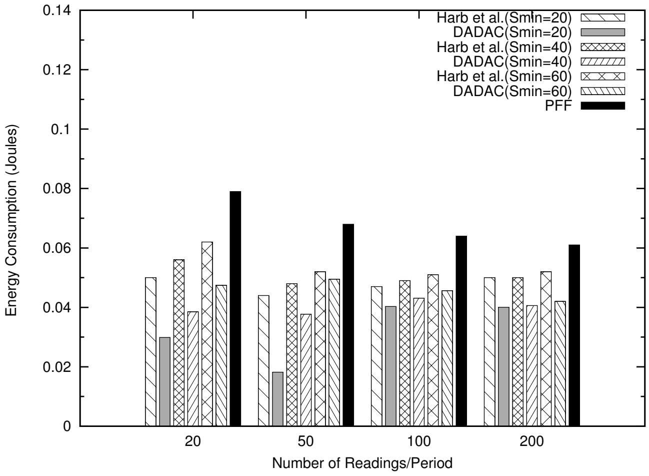



Fig. 10. Energy consumption.

As shown in Figure 10, DADAC protocol outperforms the other approaches in term of energy consumption. It saves energy because it reduces both collected and sent readings at the sensor node. The consumed energy of a sensor node using DADAC protocol is minimized from $37\%$ to $70\%$ as compared to PFF and from $20\%$ to $59\%$ as compared to Harb et al. techniques respectively. It can be observed that DADAC protocol is effective in terms of reducing energy consumption for the applications with high and low risk level, where more energy is saved when $MIN_{SMP}$ decreases.

*4) Lifetime of sensor node:* Finally, we study the influence of the number of collected and sent readings on the PSN lifetime. As exhibited by Figure 11, DADAC protocol gives a longer network lifetime compared with other approaches. Every sensor node initiated its energy to $40mJ$ for the whole approaches in this comparison.

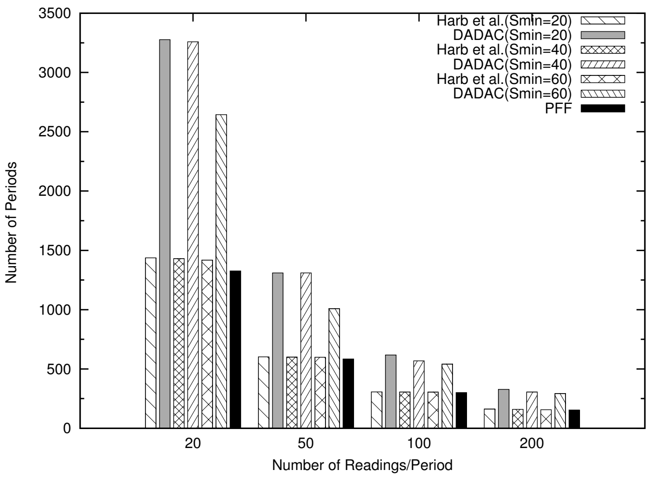



Fig. 11. Lifetime of a sensor node.

DADAC protocol enhances the lifetime of sensor node up to $56\%$ and $60\%$ comparing to the Harb et al. and PFF techniques respectively. These results are obtained due to the efficiency of DADAC protocol in conserving the energy of the sensor thus increases the PSN lifetime for both high and low risk level applications, whilst maintaining the quality of the gathered readings.

*5) Energy saving ratio:* This metric exhibits the capability of DADAC protocol to conserve energy. Figure 12 shows the energy saving ratio at the sensor node compared with PFF and Harb et al. approaches. Energy saving ratio = (1-($CE_t$/$E_{init}$))*100. $E_{init}$ refers to the initial energy of the sensor node where $E_{init} = 0.08$ joule. $CE_t$ refers to the total consumed energy at the sensor node.

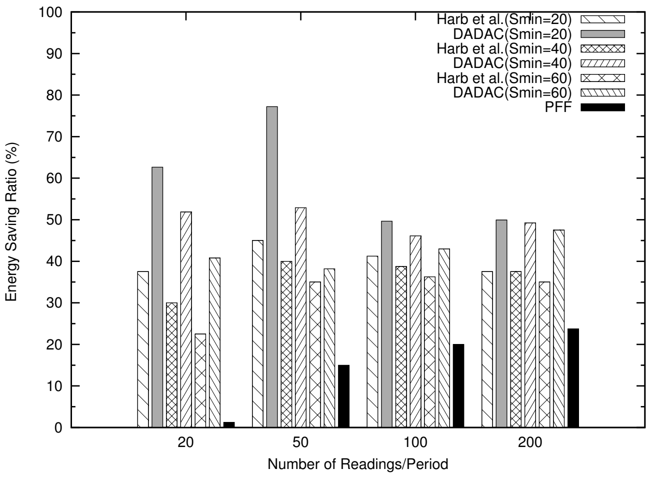



Fig. 12. Energy saving ratio.

The results show that the DADAC protocol saves efficiently energy because it reduces both gathered and transmitted readings at the sensor node. As illustrated in Figure 12, the saved energy of a sensor node using DADAC protocol is maximized from 38% to 77% whilst Harb et al. technique saved from 22% to 45%. The saved energy at the sensor node using PFF algorithm is incresed from 1% to 23%. PFF algorithm shows less energy saving compared with other approaches. However, DADAC protocol can save more energy that leads to extend the life of the sensors batteries.

*6) Algorithmic complexity:* As an analytical study, every sensor node $n_i$ constructs sequence of sensed data $S_i$ of $\rho$ temperature readings. The time complexity of the algorithm 1 is O($\rho$). The time complexity of the algorithm 2 is O($w$ $a$), where $w$ is the number of segments and $a$ is the length of Alphabet. Algorithm 3 has O($w$) as a computation complexity. The time complexity of the algorithm 4 is O(($w \cdot a$) + $\rho$). Therefore, the time complexity of our proposed DADAC protocol in the worst case is O(($w \cdot a$) + $\rho$) and it will save at most(2 $\times$ $\rho$ ) measures at the memory of the sensor node in each cycle. Therefore, the storage (space) complexity of DADAC protocol is O($\rho$). The time complexity of Harb et al. algorithm takes O($\rho^{2}$). Finally, the time complexity of PFF is O($\rho \times log_2(\rho)$). In addition, the complexity of the message in DADAC protocol depends mainly on the number of collected data ($\rho$) in the period, which is fixed by the application. If it is required a large value for $\rho$, several solutions can be used such as data packet division. The space complexity depends on the sensor node memory size as well as $\rho$, which can be handled in a similar way to the complexity of the message.

*7) The t-test:* In this section, we use the statistical analysis such as t-test to show that our results are significant. Therefore, the t-test is applied on the comparison result of the energy consumption between our proposed DADAC protocol and the two existing methods (Harb et al. and PFF). The t-test (with p-value) between DADAC and Harb et al. is equal to 0.000106089, whilst the t-test (with p-value) between DADAC and PFF is equal to 4.09444E-06. Hence, the t-test (with p-value $< 0.05$) shows that our result is significant and the energy consumption is significantly reduced.

## V. Conclusion and Future Works

This paper presents a protocol, called distributed adaptive data collection protocol (DADAC), which collects periodically sensor readings and improves the PSN lifetime. DADAC protocol works into cycles and consists of four phases. First, collecting the data readings. Second, the sensor converts the collected temperature readings into PAA representation in order to reduce its dimensionality. Third, the redundant collected readings are reduced using SAX approach. Fourth, sampling resolution to adapt the rate of sampling at the sensor node in accordance with the dynamic changing of observed environment. DADAC protocol considers the risk level of an application by fixing the minimum sampling rate that permits to sensor node to collect readings at a minimum rate while maintaining a good quality of the collected readings. To assess the effectiveness of DADAC protocol, we compared it with two other methods using several performance metrics like a number of collected and sent readings, energy consumption, and PSN lifetime. Simulation results show the efficiency of DADAC protocol to conserve the energy at the sensor nodes thus prolong the PSN lifetime.

In future, We plan to improve our work to consider the sensing overlap among sensor nodes at the aggregator level to optimize both the aggregated readings and lifetime while maintaining a good accuracy.

## Acknowledgment


The authors would like to gratefully acknowledge the University of Babylon, Iraq for the financial support.


## References


[1] E. M. Shakshuki, H. Malik, and T. Sheltami, “Wsn in cyber physical systems: Enhanced energy management routing approach using software agents,” *Future Generation Computer Systems*, vol. 31, pp. 93–104, 2014.

[2] G. Abdul-Salaam, A. H. Abdullah, M. H. Anisi, A. Gani, and A. Alelaiwi, “A comparative analysis of energy conservation approaches in hybrid wireless sensor networks data collection protocols,” *Telecommunication Systems*, vol. 61, no. 1, pp. 159–179, 2016.

[3] K. K. Chand, P. V. Bharati, and B. S. Ramanjaneyulu, “Optimized energy efficient routing protocol for life-time improvement in wireless sensor networks,” in *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*. IEEE, 2012, pp. 345–349.

[4] K. Shivanna and H. Shalini Kumari, “Energy efficient data mining in multi-feature sensor networks using improved leach communication protocol,” *IOSR Journal of Computer Engineering*, vol. 3, no. 3, pp. 8–11, 2012.

[5] A. K. Idrees, K. Deschinkel, M. Salomon, and R. Couturier, “Perimeter-based coverage optimization to improve lifetime in wireless sensor networks,” *Engineering Optimization*, vol. 48, no. 11, pp. 1951–1972, 2016.

[6] A. Idrees, K. Deschinkel, M. Salomon, and R. Couturier, “Distributed lifetime coverage optimization protocol in wireless sensor networks,” *The Journal of Supercomputing*, vol. 71, no. 12, pp. 4578–4593, 2015.

[7] A. K. Idrees, K. Deschinkel, M. Salomon, and R. Couturier, “Coverage and lifetime optimization in heterogeneous energy wireless sensor networks,” *ICN 2014*, p. 60, 2014.

[8] C. Alippi, G. Anastasi, C. Galperti, F. Mancini, and M. Roveri, "Adaptive sampling for energy conservation in wireless sensor networks for snow monitoring applications," in *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*. IEEE, 2007, pp. 1–6.

[9] X. Tang and J. Xu, "Adaptive data collection strategies for lifetime-constrained wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 6, pp. 721–734, 2008.

[10] A. K. I. Ali K. M. Al Qurabat, "Distributed data aggregation protocol for improving lifetime of wireless sensor networks," *The 1st International Conference on Information Technology (ICoIT'17)*, pp. 220–230, 2017.

[11] A. K. I. Wesam Huseein, "Energy - efficient sensor activity scheduling protocol for wireless sensor networks," *The 1st International Conference on Information Technology (ICoIT'17)*, pp. 257–265, 2017.

[12] G. Campobello, A. Segreto, and S. Serrano, "Data gathering techniques for wireless sensor networks: A comparison," *International Journal of Distributed Sensor Networks*, vol. 2016, 2016.

[13] L. A. Tang, J. Han, and G. Jiang, "Mining sensor data in cyber-physical systems," *singhua Science and Technology*, vol. 19, no. 3, pp. 225–234, 2014.

[14] A. Makhoul, H. Harb, and D. Laiymani, "Residual energy-based adaptive data collection approach for periodic sensor networks," *Ad Hoc Networks*, vol. 35, pp. 149–160, 2015.

[15] A. K. I. Ali K. M. Al Qurabat, "Adaptive data collection protocol for extending lifetime of periodic sensor networks," *The 1st International Conference on Information Technology (ICoIT'17)*, pp. 108–118, 2017.

[16] J. M. Bahi, A. Makhoul, and M. Medlej, "A two tiers data aggregation scheme for periodic sensor networks." *Adhoc & Sensor Wireless Networks*, vol. 21, no. 1, 2014.

[17] H. Harb, A. Makhoul, A. Jaber, R. Tawil, and O. Bazzi, "Adaptive data collection approach based on sets similarity function for saving energy in periodic sensor networks," *International Journal of Information Technology and Management*, pp. 1–19, In Press 2016. [Online]. Available: http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijitm.

[18] S. Chatterjea and P. Havinga, "An adaptive and autonomous sensor sampling frequency control scheme for energy-efficient data acquisition in wireless sensor networks," in *International Conference on Distributed Computing in Sensor Systems*. Springer, 2008, pp. 60–78.

[19] A. Masoum, N. Meratnia, and P. J. Havinga, "A decentralized quality aware adaptive sampling strategy in wireless sensor networks," in *Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference on*. IEEE, 2012, pp. 298–305.

[20] R. Willett, A. Martin, and R. Nowak, "Backcasting: adaptive sampling for sensor networks," in *Proceedings of the 3rd international symposium on Information processing in sensor networks*. ACM, 2004, pp. 124–133.

[21] J. Wang, S. Tang, B. Yin, and X.-Y. Li, "Data gathering in wireless sensor networks through intelligent compressive sensing," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 603–611.

[22] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 7, pp. 1010–1023, 2007.

[23] B. Gedik, L. Liu, and S. Y. Philip, "Asap: an adaptive sampling approach to data collection in sensor networks," *IEEE Transactions on Parallel and distributed systems*, vol. 18, no. 12, pp. 1766–1783, 2007.

[24] A. Masoum, N. Meratnia, and P. J. Havinga, "An energy-efficient adaptive sampling scheme for wireless sensor networks," in *Intelligent Sensors, Sensor Networks and Information Processing, 2013 IEEE Eighth International Conference on*. IEEE, 2013, pp. 231–236.

[25] F. Alassery and W. K. Ahmed, "Toward energy efficient wireless sensor networks using pilot periods for collision detection," *Engineering Letters*, vol. 24, no. 1, pp. 113–124, 2016.

[26] F. Alassery, W. K. Ahmed, M. Sarraf, and V. Lawrence, "A low computational complexity statistical discrimination algorithm for collision detection in wireless sensor networks," *IAENG International Journal of Computer Science*, vol. 41, no. 3, pp. 204–211, 2014.

[27] Y. MALEH and A. Ezzati, "Lightweight intrusion detection scheme for wireless sensor networks," *IAENG International Journal of Computer Science*, vol. 42, no. 4, pp. 347–354, 2015.

[28] W.-H. Kuo, Y.-S. Chen, K.-T. Cheng, and T.-W. Lu, "Signal strength based indoor and outdoor localization scheme in zigbee sensor networks." *IAENG International Journal of Computer Science*, vol. 23, no. 1, pp. 55–64, 2016.

[29] P. Chaturvedi and A. Daniel, "A hybrid scheduling protocol for target coverage based on trust evaluation for wireless sensor networks." *IAENG International Journal of Computer Science*, vol. 44, no. 1, pp. 87–104, 2017.

[30] C. Liu, K. Wu, and M. Tsao, "Energy efficient information collection with the arima model in wireless sensor networks," in *GLOBECOM'05. IEEE Global Telecommunications Conference, 2005.*, vol. 5. IEEE, 2005, pp. 5–pp.

[31] Y. W. Law, S. Chatterjea, J. Jin, T. Hanselmann, and M. Palaniswami, "Energy-efficient data acquisition by adaptive sampling for wireless sensor networks," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*. ACM, 2009, pp. 1146–1151.

[32] P. Padhy, R. K. Dash, K. Martinez, and N. R. Jennings, "A utility-based adaptive sensing and multihop communication protocol for wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 3, p. 27, 2010.

[33] A. Jain and E. Y. Chang, "Adaptive sampling for sensor networks," in *Proceeedings of the 1st international workshop on Data management for sensor networks: in conjunction with VLDB 2004*. ACM, 2004, pp. 10–16.

[34] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 2, pp. 335–344, 2010.

[35] D. Laiymani and A. Makhoul, "Adaptive data collection approach for periodic sensor networks," in *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 2013, pp. 1448–1453.

[36] B. Srbinovski, M. Magno, B. O'Flynn, V. Pakrashi, and E. Popovici, "Energy aware adaptive sampling algorithm for energy harvesting wireless sensor networks," in *Sensors Applications Symposium (SAS), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[37] J. Zhang, L. Ren, Y. Ding, and K. Hao, "Adaptive sampling algorithm with endocrine regulation mechanism for wireless sensor network," in *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*. IEEE, 2015, pp. 502–507.

[38] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, "Similarity measures and dimensionality reduction techniques for time series data mining," *Advances in data mining knowledge discovery and applications*, 2012.

[39] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, pp. 2–11.

[40] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[41] Wikipedia, "Sorting — wikipedia, the free encyclopedia," 2016, [Online; accessed 23-August-2016]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sorting&oldid=733029871

[42] J. Lin and Y. Li, "Finding structural similarity in time series data using bag-of-patterns representation," in *International Conference on Scientific and Statistical Database Management*. Springer, 2009, pp. 461–477.

[43] S. Malinowski, T. Guyet, R. Quiniou, and R. Tavenard, "1d-sax: A novel symbolic representation for time series," in *International Symposium on Intelligent Data Analysis*. Springer, 2013, pp. 273–284.

[44] A. Varga's, "Omnet++ discrete event simulator." 2003. [Online]. Available: https://omnetpp.org/.

[45] S. Madden, "Intel berkeley research lab," 2004, [Online; accessed 25-August-2016]. [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html

**Ali Kadhum IDREES** received his B.Sc. and M.Sc. degrees in Computer Science from the University of Babylon, Iraq in 2000 and 2003 respectively. He received his Ph.D. degree in Computer Science (Wireless Networks) in 2015 from the University of Franche-Comte (UFC), France. He is currently an assistant professor in Computer Science at the University of Babylon, Iraq. His research interests include Wireless Networks, WSNs, SDN, IoT, Distributed Computing, Data Mining, and Optimization in Communication Networks.

**Ali Kadhum M. Al-Qurabat** is currently a Ph.D. student in the Software Department, University of Babylon, Iraq. He received the M.Sc. degree in IT from Universiti Tenaga Nasional (UNITEN), Malaysia, in 2012. He received the B.Sc. degree in computer science from University of Babylon, Iraq, in 2002. He joined the Department of Computer Science, University of Babylon, Iraq, in 2006, where he is currently an assistant Lecturer. His research interests include E-Procurement, sensor networks, WSN, data aggregation.