Journal of Engineering and Applied Sciences 13 (Special Issue 5): 4644-4653, 2018
ISSN: 1816-949X

Distributed Data Aggregation and Selective Forwarding Protocol for Improving Lifetime of Wireless Sensor Networks

[1]Ali Kadhum M. Al-Qurabat and [2]Ali Kadhum Idrees
[1]Department of Software, College of Information Technology,
[2]Department of Computer Science, College of Sciences for Girls,
University of Babylon, Babylon, Iraq

Abstract: In this study, a Distributed Data Aggregation and Selective Forwarding (DiDASeF) protocol for prolonging the lifetime of wireless sensor networks is suggested. DiDASeF combines two energy efficient approaches for a clustered network: data aggregation and selective forwarding. DiDASeF works into cycles and each cycle consists of two stages. In the first stage, DiDASeF aggregates and reduces data dimensionality by using an Adaptive Piecewise Constant Approximation (APCA) method. The selective forwarding by using Dynamic Time Warping (DTW) distance measure is implemented in the second stage. DiDASeF was successfully evaluated using OMNeT++ network simulator and based on sensed data of a real sensor network. The conducted simulation results show that the proposed DiDASeF protocol decreases the consumed energy and extending the network lifetime in comparison with some existing methods whilst keeping the sensed data quality at the sink node.

Key words: Wireless sensor networks, data aggregation, selective forwarding, APCA and DTW, network lifetime, DiDASeF

INTRODUCTION

A WSN consists of a large number of tiny low-cost limited-energy devices that can sense, process, store and transmit data of surrounding environment with limited capabilities across the network to the sink node. The most significant resource in the sensor node that impacts on the lifetime of WSN is the energy provided by the battery. Since, the limited lifetime of the battery in the sensor node, it is difficult or impossible to replace (or recharge) it especially in the remote or hostile environment. Therefore, the lifetime maximization of the battery represents one of the biggest challenges in WSN (Idrees *et al*., 2015, 2016).

Since,s the limited power of sensor nodes batteries, it is important to transmit as less volume of data as possible to the base station (sink). The sensed data received by the sink node may be similar because there is more than one sensor monitors the same region. Therefore, it is necessary to improve the energy efficiency of the sensor network to operate over a long period of time (Idrees *et al*., 2015, 2016; Anastasi *et al*., 2009; Zhai and Vladimirova, 2015). In the literature, several energy-saving strategies are applied to the sensor networks such as scheduling, routing, clustering, battery repletion, radio optimization and data-driven approaches (Povedano *et al*., 2014). Data aggregation methods are efficiently applied to WSN in order to remove the redundant data and decrease the communication cost, thus, enhance the network lifetime (Dalbro *et al*., 2008).

Based on application requirements, data gathering can be either triggered events (such as forest fire and gas or oil leaks detection (Mainwaring *et al*., 2002) or periodic triggering (such as habitat monitoring (Bahi *et al*., 2014).

This study focuses on the periodic data gathering and aggregation in WSNs. In some specific WSN applications, the accuracy of the observations is very critical for understanding the underlying processes. Therefore, in order to design data aggregation algorithms for such applications, it is very important to ensure the accuracy of the received sensed data by the sink node.

This study provides the following contributions. A new protocol named DiDASeF (Distributed Data Aggregation and Selective Forwarding) is proposed to aggregate the sensed data and prolong the network lifetime in WSNs. It uses an energy efficient method for data aggregation and selective forwarding for a clustered network. Selective forwarding method is proposed that

Corresponding Author: Ali Kadhum M. Al-Qurabat, Department of Software, College of Information Technology, University of Babylon, Babylon, Iraq

focuses on the transmission function of the sensor nodes, especially the nodes of the same cluster. Instead of sending sensed data immediately to the cluster head, a cluster member must pass the transmission criteria.

DiDASeF protocol is evaluated by OMNeT++ network simulator using extensive simulation experiments. DiDASeF has been compared to two algorithms in the related works: PFF algorithm that proposed by Harb *et al.* (2016) and ATP protocol proposed by Sharaf *et al.* (2003).

Literature review : This study investigated some existing related works to data aggregation in WSNs. The principle objective of aggregating the sensed data is to eliminate the redundancy in the sensed data and minimize the consumed energy, thus, extending the lifetime of the network (Ren *et al*., 2013).

However, in order to reduce the data transferring, the researchers by Xu *et al.* (2012) and Heinzelman *et al.* (2000) used simple aggregation methods (such as MAX, MIN, AVG and SUM) for aggregation. These methods do not consider the correlation among the sensed data. Although, they provided a high aggregation performance but the accuracy of recovered data is badly poor. Therefore, these methods are inappropriate for those applications that require a high data accuracy. For example, LEACH (Villas *et al.*, 2014) protocol was divides the network into several clusters. The cluster heads are chosen during the setup phase whilst the data are aggregated using AVG method at each cluster head in the steady phase so as to reduce the network data traffic.

The reseach by Tran and Oh (2014) divided the sensor nodes into different clusters. One or several nodes in each cluster are chosen as a representative set of nodes for data collecting and sending whilst deactivating the other nodes in the same cluster. The node's energy can be saved significantly with these methods but this can lead to important data loss due to a large number of deactivated nodes.

The reseach by Chong *et al.* (2007) proposed a round-based clustering scheme that resolves the transmission of redundant data in the network, so as to improve network lifetime. Proposed scheme works in four phase's rounds: initialization, cluster head selection, clustering and data aggregation. Proposed clustering scheme reduces energy consumption, thus, increasing network throughput by dealing with most of the redundant data.

MATERIALS AND METHODS

Description of the DiDASeF protocol: The description of DiDASeF protocol is given in more details in this study. The primary goal of DiDASeF is to develop a cluster based data aggregation and selective forwarding protocol that works at the sensor node level. DiDASeF protocol consists of two stages. The aim of the first stage is to apply an energy efficient data aggregation inside each sensor node before sending the aggregated data to the appropriate Cluster Head (CH). In the second stage, selective forwarding makes each sensor node compare the data of successive periods in order to decide send this data or not to CH based on the amount of similarity between them. As a result, the redundancy in the collected readings will be reduced and the consumption of energy will be minimized (WSN lifetime improvement) while the quality of collected readings is kept up adequately to permit for a significant analysis at the base station. Figure 1 illustrates the flowchart of the proposed DiDASeF protocol. Table 1 explains some parameters used in this study.

Table 1: Some parameters used in this study

SMP_F	Sampling rate = p
S	Temperature readings series $S = s_1, ..., s_n$
S^W	Segment construction using $SW(S, \epsilon)$, $S^W = s^W_1, ..., s_m{}^W$
S^{AP}	APCA of S^W, $S^{AP} = c_1{}^{ap}, ..., c_w{}^{ap}$
ε	Reconstruction error bound
n	Sensor id
n_e	Remaining energy of sensor n

Data aggregation stage using APCA: The proposed protocol in this article is distributed on the sensor nodes. These nodes are considered grouped into clusters so as to achieve energy efficient data aggregation with reduced cost of communication. DiDASeF protocol is a periodic and works into periods. The PSN consists of N nodes (n_1, n_2, ..., n_N) each node is responsible for sensing the data measures of the dynamic physical environment such as humidity, temperature or pressure, etc., In PSN, the periods are partitioned into time slots. Therefore, each sensor node n captures the data reading periodically.

Consequently, the time-ordered sequence of sensed data constitutes a time series $S_i = \{s_1, s_2, \ldots, s_{p-1}, s_p\}$ where ρ is the total number of temperature readings generated by sensor node n_i every T sec. Therefore, DiDASeF protocol treats the sensor readings as a time series and named it as a temperature readings series. 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.

The dimensionality ρ of temperature readings series (which is the number of observed measures) have a direct proportionality relation with the communication cost (Fig. 1).

Thus, a smaller ρ can result in a significant reduction in the communication cost and hence, it will prolong the lifetime of the sensor network (Wang *et al.*, 2013). In this stage, DiDASeF protocol transforms the temperature readings series S that collected during the period to an APCA representation in order to decrease Fig. 1

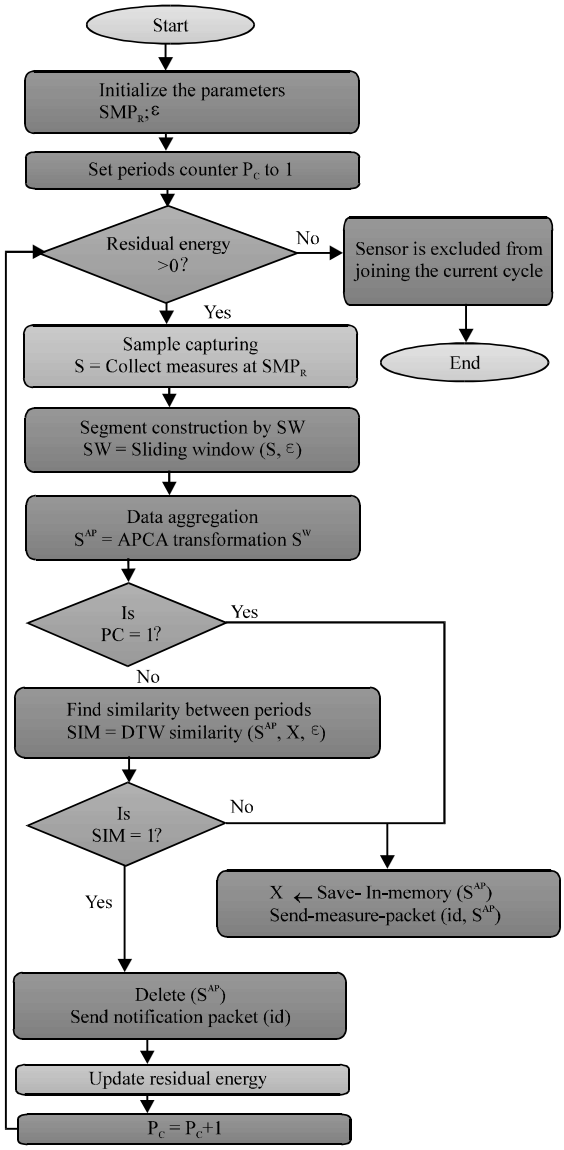

Fig. 1: Flowchart of proposed DiDASeF protocol

dimensionality of series. It exploits the correlation nature which is temporal among the sensed data of the sensor node efficiently by applying an Adaptive Piecewise Constant Approximation (APCA) technique. The efficiency of APCA is improved by sorting the sensed temperature readings in descending order, so as to group the similar (or close similar) readings together.

The APCA divides the sorted temperature readings series S into a set of constant value segments (with a bounded reconstruction error ε) of varying lengths based on data such that their individual reconstruction errors are minimal. More formally, $| R(S^{AP}) - S | < \varepsilon$, $R(S^{AP})$ is the reconstruction function and ε is an error threshold. Long segments are used to represent data regions of low activity and short segments are used to represent regions of high activity (Yahmed *et al* 2015). The APCA representation of S is given in Eq. 1:

$$S^{AP} = \{(dm_1, dr_1), ..., (dm_m, dr_m)\}, dr_0 = \quad (1)$$

The APCA approximates each segment S_j^{AP} by a pair (dm_j, dr_j) of two numbers where dm_j is the mean value of temperature readings in the jth segment. Whilst dr_j is the right endpoint of the jth segment (Yahmed *et al*., 2015).

By using the standard form of APCA with a constant number of segments of varying lengths can influence on the accuracy of temperature readings. Hence, the problem addressed here is: for a given temperature readings series S and a given reconstruction error bound ε, find the number of segments to approximate the time series such that the difference between any approximation value and its actual value is <ε. In our method we make some slight modifications on APCA. First, the number of segments m will not be constant and predetermined but it will be adaptive based on the user specified reconstruction error ε. In order to achieve this goal (i.e., making the number of segments adaptive), the sliding window algorithm is utilized. The reason for making the number of segments adaptive is to increase the accuracy of approximated measures by using a user-specified reconstruction error. Second, we modified d_r to represent the length of the segments rather than record the locations of their right endpoints.

Sliding window algorithm: Several applications such as weather, medical and stocks employ the algorithm of the sliding window. It is a temporal approximation over the actual value of the time series data (Gedik *et al*., 2007). At the end of each period, DiDASeF protocol will apply the sliding window algorithm on the collected readings to produce a different number of segments with varying lengths.

The sliding Window approach is used because it is simple, online and intuitive (Gedik *et al*., 2007). Algorithm 1 represents the process of segment construction using sliding window algorithm.

Algorithm 1; Segments construction by sliding window:

Input: ε Reconstruction Error bound
S (ρ-dimensions temperature readings series)
Output: S^W the set of segments with m subsets
Process:
1: S←Sorting (S) //Sorting temperature readings
2: FLAG←1 // Starting point
3: SEG_{No}←1 // Number of Segments
4: while (x <ρ) do
5: x←2
6: while (Calculate-Error (S [Flag: Flag+x]) <ε) do

7: x←x+1
8: end while
9: S^W [SEG]←Create-Segment (S [Flag: Flag+x-1]
10: Flag←Flag +x
11: SEG_{No}←SEG_{No}+1
12: end while
13: return S^W

Let S_i^W be the subset consisting of all the temperature readings on this segment $\{s_i, s_{i+1}, \ldots, s_j\}$ which meet the reconstruction error bound. Eventually, we have a set S^W of m subsets where $S^W=(S_1^W, S_2^W, \ldots, S_m^W)$. After segmenting the temperature readings series using sliding window algorithm, the produced set of segments S^W is used by Algorithm 2 to produce the APCA representation for temperature readings series S. Algorithm 2 illustrates the process of dimensionality reduction using APCA. Let $S_i^{AP}(SEG_{\mu i}, SEG_{leni})$ denote a subset consisting of all the temperature readings on this segment $\{s_i, s_{i+1},\ldots, s_j\}$, where $SEG_{\mu i}$ is the mean of these temperature readings and SEG_{leni} is the length of the segment. The problem mentioned above is solved by constructing a set of segments S^{AP} with m subsets that meet the reconstruction error bound ε.

Algorithm 2; Aggregation stage using APCA:

Input: S^W the set of segments with m subsets
Output: S^{AP} the set of segments with m subsets and two numbers per segment
Process:
1: for I←1 to m do
2: SG←S_i^W
3: Sum←0
4: Count←0
5: For j←1 to Len (SG) do
6: Sum←Sum+SG [j]
7: Count+1
8: end for
9: SEG_{len}←Count
10: SEG_{μ}←sum/count
11: S_i^{AP}←create-segment (SEG_{μ} SEG_{len})
12: end for
13: return S^{AP}

In the aggregation stage and at the end of every period, each sensor n_i will have a set of segments S^{AP} with m subsets ($S_1^{AP}, S_2^{AP}, \ldots, S_m^{AP}$)and two number per segment ($SEG_{\mu i}$, SEG_{leni}) that meet the reconstruction error bound ε with no redundant measures. Then, each sensor runs the selective forwarding in the second phase as explained in the next section.

Selective forwarding stage: In WSN, the strong factor of energy consumption is the radio communication. Therefore, sending a lot of data to the base station can assist in various undesired issues such as increasing communication overhead, network congestion and energy consumption (Harb *et al*., 2014). At the end of the first stage, the data reading set in each sensor node n_i is constructed. The node n_i provides a decision about transmitting or not this data set to the cluster head. The similarity between the successive periods is calculated by the sensor node in this stage in order to adapt data set transmission to the cluster head. If the two successive data sets are similar, the sensor node does not transmit the current period data set, instead, it will send a notification packet to the cluster head inform him that the current data set is similar to the previous period data set. This can save the energy of the sensor node and decrease the communication cost thus increase the PSN lifetime. The data model of the PSN is applied for monitoring the dynamic changing environment in which the physical phenomena can change fastly or slowly (Cassisi *et al*., 2012). In the fast changing environment (or when the period p is short), the node transmits a higher redundant data to the cluster head. Therefore, the sensor node must adapt its data set forwarding to this dynamic changing of the monitored environment in order to reduce the transmitted data set to the cluster head and improve the PSN lifetime.

Similarity measure: The main purpose of using similarity measure is to exploit the similarity among periods in order to decide transfer or not the sensed data to appropriate cluster head according to the amount of similarity among successive periods. The data aggregation stage at the end of each period in DiDASeF protocol will use the modified APCA technique on the collected measures and produce for each period a different number of segments with varying lengths. Since, it is not suitable to use the Euclidean distance to calculate the distance between sequences whose lengths are different. Therefore, Dynamic Time Warping (DTW) distance method has been adopted to overcome these problems. It is an important data mining measure that used in several time series problems such as classification, clustering and anomaly detection, that allows time-axis scaling. DTW is able to measure the distance between two temperature readings series of varying lengths. It is not used one-to-one comparison such as in Euclidean but it uses many-to-one (and vice versa) comparison. If we have two temperature readings series $Q = (q_1, q_2, \ldots, q_p)$ and $T = (t_1, t_2, \ldots, t_m)$ of length p and m, respectively, we will build an p-by-m distance matrix in order to align the two sequences by minimizing the sum of squared Euclidean distances using DTW.

Where the element in the position (i^{th}, j^{th}) of the matrix contains distance $d(q_i, t_j)$ between q_i and t_j. Usually distance used in this matrix between two points is squared Euclidean distance. Each matrix element (i, j) corresponds

to the alignment between the points q_i and t_j. The goal of DTW is to find the warping path W = {w1, w2, ..., wk, ..., wK} of adjacent elements on DistMtrx where (max (p, m) = K<p + m -1 and wk = DistMtrx (I, j) such that it minimizes the following function in Eq. 2:

$$DTW(Q, T) = \min\left\{\sum_{K=1}^{K}(w_k)/K\right\} \quad (2)$$

The warping path is ordinarily subject to few restrictions (Cassisi *et al.,* 2012). Given w_k = (I, j) and w_{k-1} = (i', j') with i, i'≤p and j, j'≤m:

- Boundary conditions: w1 = (1,1) and wK = (p, m)
- Continuity: i – i'≤1 and j – j≤1
- Monotonicity: i – i'≥0 and j – j'≥0

The dynamic programming is used to find this path to assess this recurrence that defines the cumulative distance matrix ε(I, j) of the same dimension as the DistMtrx where the distance d (i,j) is found in the current cell and the minimum of the cumulative distances of the adjacent elements Eq. 3:

$$\gamma(i, j) = d(q_i, t_j) + \min\{\gamma(i\text{-}1, j\text{-}1), \gamma(i\text{-}1, j), \gamma(i, j\text{-}1)\} \quad (3)$$

The w_K refer to the last warping path element that meets the computed distance with the DTW approach (Madden). After finished the distance calculation between two temperature readings series that transformed to APCA representation Q_p and T_m, our protocol uses a similar function to identify the similarity between them. The function returns one of two values 0 or 1. If the value is 0, then temperature readings series are entirely different, while if the value is 1, then it means that temperature readings series are similar. The similar function refers to the similarity between two APCA temperature readings series using the following formula Eq. 4:

$$SIM(Q_p, T_m) = \begin{cases} 1 & \text{if } \|Q_p - T_m\| \le \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where, ε is a reconstruction error fixed by the application. Furthermore, two temperature readings series are similar if and only if their similar function is equal to 1. Algorithm 3 illustrates the similarity calculation between two APCA temperature readings series Q_p and T_m.

Algorithm 3; Similarity algorithm:

Input: ε: Reconstruction Error bound
Two APCA temperature series Q_p and T_m
Output: Sim
Process:

```
for i←1 to len (Q) do
  for j ← 1 to len (T) do
  Distance [i, j] ← (Q [I]-T[j])^2
  end for
end for
Accumulated-Cost [1, 1]←Distance [1, 1]
for I←1 to len (Q) do
   Accumulated-Cost [i, 1]←Distance [I, 1]+ Accumulated-Cost [i, 1, 1]
end for
for j←1 to len (T) do
   Accumulated-Cost [1, j]←Distance [1, j]+Accumulated-Cost [1, j-1]
end for
for i←1 to len (Q) do
  for j←1 to len (T) do
  Accumulated-Cost [i, j]←Distance [i, j]+min (Accumulated-Cost [i, j, 1] Accumulated-Cost [i-1, j-1])
   end for
end for
if (Accumulated- Cost [p, m] ≤ε) then
   Sim←1
else
   Sim←0
end if
return Sim
```

Selective forwarding approach: In the second stage, the selective forwarding approach is applied, so, as to adapt the data set forwarding in each sensor node to the cluster head. This data transmission adaptation can contribute to conserving the energy and improving the PSN lifetime. Two types of packets are used by the sensor node in this stage: data packet or control packet. The latter is an empty control packet that used to notify the cluster head that the current data set is similar to the previously transmitted data set. The former includes the data of the constructed data set in the current period. As soon as the first stage is finished in this period, the selective forwarding approach calculates the similarity between the newly constructed data set and the previously transmitted data set. After that, it decides to send a control-packet or data packet to the cluster head based on the percentage of the similarity between the two data sets. A control packet is sent to the cluster head when the two data sets are similar to prevent data repetition and decrease the consumed energy. Otherwise, the sensor node saves the current data set and send it by a data-packet to the cluster head. The received data set of each node will be stored in the memory of the cluster head. The cluster head updates its memory periodically. It transmits all the received data sets of the sensor node to the sink at the end of each period. The sink node sends the whole data set to the end user for further analysis. The idea in this stage is inspired from (Cassisi *et al.,* 2012) with some modifications. DiDASeF protocol applied a selective forwarding approach for data set transmission adaptation based on the Dynamic Time Warping (DTW) distance to prevent the node from

sending the current sensed data which have high correlations with the previously sensed data to the cluster head as illustrated in algorithm 4.

Algorithm 4; Selective-forwarding algorithm:

Input: Sensor id, set of measures S_i^{AP} at perid j
saved set measures S_p^{AP}
ε: Reconstruction Error
Output: saved set of measures S_p^{AP}
Process:
1: if is the fisrt period then
2: Save-in-memory←S_i^{AP}
3: Send-Measure-Packet (id, S_i^{AP})
4: else
5: if Similarity (S_i^{AP}, S_p^{AP}, ε) = 1 then
6: Delete (S_i^{AP})
7: Send-Similarity-notification (id)
8: else
9: Save-in-memory←ø
10: Save-in-memory←S_i^{AP}
11: Send-Measure-Packet (id, S_i^{AP})
12: end if
13: end if
14: return

RESULTS AND DISCUSSION

Protocol evaluation

Simulation framework: In order to evaluate DiDASeF protocol, extensive simulations experiments are performed with discrete event simulator OMNeT++ and based on real sensor data. In these simulations, we consider N sensors deployed in the Intel Berkeley Research Lab, (Harb *et al.*, 2014). Sensors periodically capture local readings (e.g., temperature) at a specified rate. We assume there is a single cluster head located at the center of the lab. The cluster head receives sensed data readings from each sensor node in the lab periodically via a single hop. DiDASeF protocol is distributed at each sensor node and it is based on the dataset of Intel Berkeley Research Lab, (Harb *et al.*, 2014). 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 sec. 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 study uses only one measure of sensor node measurements: temperature. There are seven sensor nodes not used in our simulation because its data may be missed or truncated. Therefore, the results are the average of 47 sensor nodes. Table 2 gives the selected parameters settings.

In the experimental simulations, some performance metrics are applied to assess the effectiveness of the DiDASeF protocol such as percentage of data after aggregation, the percentage of data sets sent to the cluster head, data accuracy and energy consumption. DiDASeF protocol uses the same energy consumption model discussed in Technology and Management.

Table 2: Simulation parameters for PSN initialization

Parameters	Values
PSN size	47 nodes
ρ	20, 50 and 100 readings
ε	0.03, 0.05 and 0.07
E_{elec}	50 nJ/bit
β_{amp}	100 pJ/bit/m^2

Energy consumed by the sensor node is caused by the communication unit (data transmission and reception). Therefore, the cost of transmission is calculated for a m-bits message and for a distance d as in Eq. 5:

$$E_{TX}(m,d)=E_{elec}*m+\beta_{amp}*m*d^2 \quad (5)$$

The energy consumption required for reception m-bits is calculated as in Eq. 6:

$$E_{RX}(d,m)=E_{elec}*m \quad (6)$$

These experiment simulations consider the length of data reading m equal to 64. In the case of transmission, 64 bits are added to m bits message which corresponds to the frequency of data reading m. The length of the transmitted data packet is calculated as follow data packet length = (number of readings in the data set ×2) ×64 bits. Hence, the packet length is the number of reading in the sensed data set with their frequencies multiplying by 64 bits.

Performance comparison and analysis: Several experiments are achieved in this section to show the performance of DiDASeF protocol. DiDASeF is distributed at each sensor node in the PSN. Every node reads real temperature readings periodically and aggregate them in the first stage and then applied selective forwarding algorithm in the second stage to decide whether send or not the current temperature readings based on the similarity percentage among collected sets of temperature readings. Furthermore, DiDASeF protocol is compared to two existing data aggregation approaches: ATP (Harb *et al.*, 2014) and PFF (Harb *et al.*, 2015). ATP method works into two steps: in the first step, ATP calculates the similarities between collected data to remove the duplicated data. The second step calculates the similarity between the data using one-way ANOVA model and Fisher to decrease the number of transmitted data sets to the cluster head. Harb *et al.* (2015) two level data aggregation is performed. The first level achieves a local processing inside the node, whilst Jaccard similarity method is used by PFF at the aggregators level to combine the sensed data after removing the similar data that received by the close nodes. For simplicity, the parameter δ in the next figures is equivalent to reconstruction error bound ε.

Percentage of data after applying aggregation stage: The result of the aggregation in this stage depends on the chosen reconstruction error bound ε (δ), the number of the collected measures ρ in the period and the changes in the monitored region. Figure 2 illustrates the remaining data Percentage without and with applying aggregation stage at the end of simulation by every sensor node using DiDASeF protocol compared with ATP and PFF approaches. The results show a maximum of 10% of the data remains after applying the aggregation stage by DiDASeF protocol at each period, whilst the rate is equal to 31% after applying the aggregation step in ATP and 100% without applying the aggregation step in PFF. Therefore, DiDASeF protocol decreases the volume of sensed data transmitted to the cluster head by removing the duplicated measures at every period successfully. It can be seen at the step of aggregation when the ρ or ε(δ) increases, the redundant data are increases. The reason behind this is to remove a larger amount of similar data by using APCA method.

Percentage of transmitted sets to the cluster head at the second stage: In this experiment, the transmitted sets percentage by the sensor node to the cluster head after implementing the selective forwarding stage is investigated. Figure 3 displays the transmitted sets percentage by the sensor node and for three protocols (DiDASeF, ATP and PFF). By using various values for both ñ and δ(ε), the comparison results illustrate the reduction in the transmitted data sets to the cluster head from 15- 63% using the proposed DiDASeF protocol while ATP decreases the transmitted data sets to cluster head from 10-17%. The percentage of transmitted data sets by the node to the cluster head using PFF is 100% because it does not use adaptive transmission inside the sensor node. Hence, the selective forwarding stage allows to each sensor node to adapt its transmitted sensed data according to the real modifications in the monitored environment.

Therefore, our protocol outperforms the two techniques: ATP and PFF where it is successfully, achieved data sets reduction sent to its proper cluster head at each period. We can also observe that at the selective forwarding stage, the sensor node transmits a higher data sets when δ(ε) decreases. In addition, the sensor node removes a larger number of data sets when ρ decreases because of the high similarity between gathered data in the short periods. Therefore, selective forwarding stage becomes more efficient when ρ decreases.

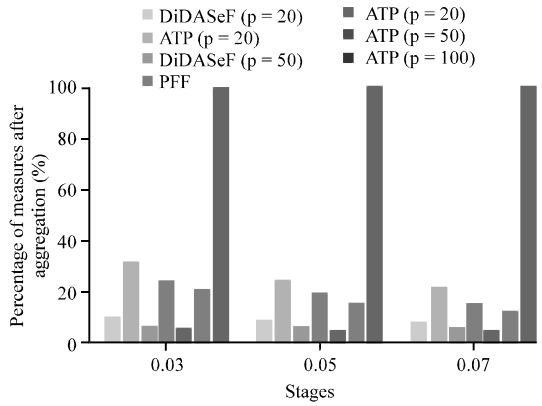

Fig. 2: Percentage of data after aggregation stage

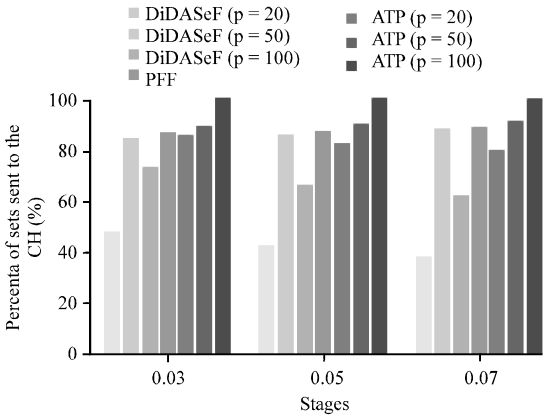

Fig. 3: Percentage of sets sent to CH

Data accuracy: In this experiment, the data accuracy is considered as an essential performance factor in WSNs. In this study, it represents the percentage of data loss after applying the aggregation and selective forwarding inside the sensor node. It has a significant impact on the final decision that will be taken by the end user. It can be considered as the error of aggregation. In this section, the proposed DiDASeF protocol is compared with the ATP and PFF techniques. Figure 4 shows the results of data accuracy for our technique DiDASeF, ATP and the PFF.

It can be seen that our protocol provides better results from the data accuracy point of view. DiDASeF outperforms ATP and PFF in all cases. In the worst case, the percentage of data which are not received by the sink are almost 1.06% (i.e., δ (ε) = 0.03 and ρ = 20 in Fig. 4a. This percentage is not important in comparison with the received data by the base station. Therefore, it can be noted that our protocol is able to get rid of the redundant data while maintaining the accuracy of received data by the end user. Furthermore, the data loss percentage

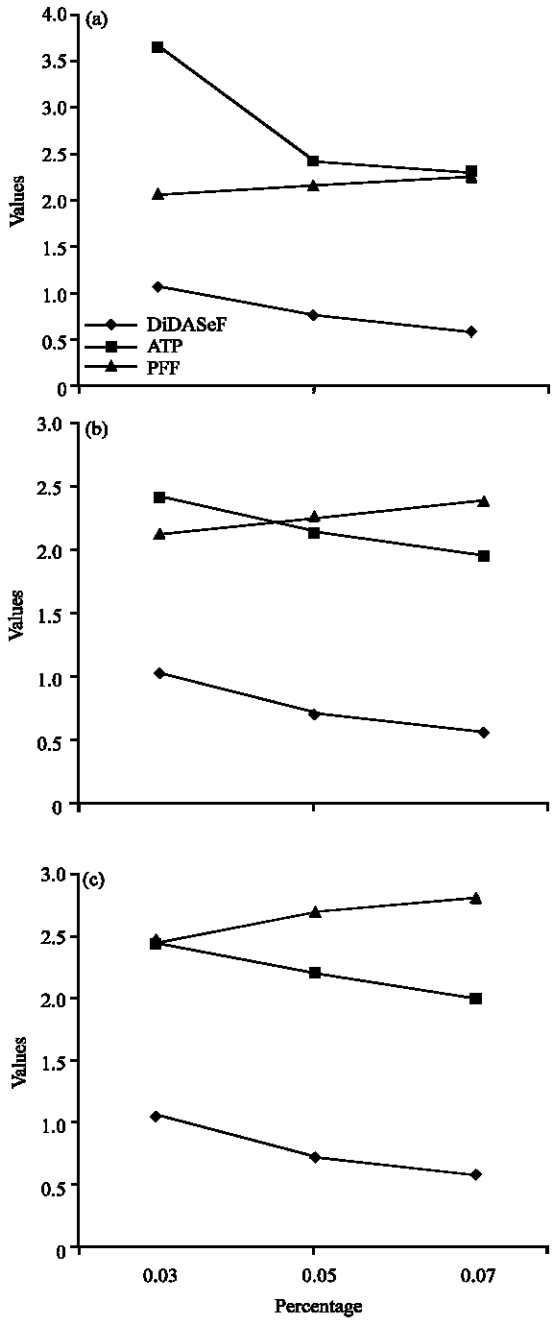

Fig. 4: Data accuracy (percentage of lost measures (%): a) p = 20; b) p = 50 and c) p = 100

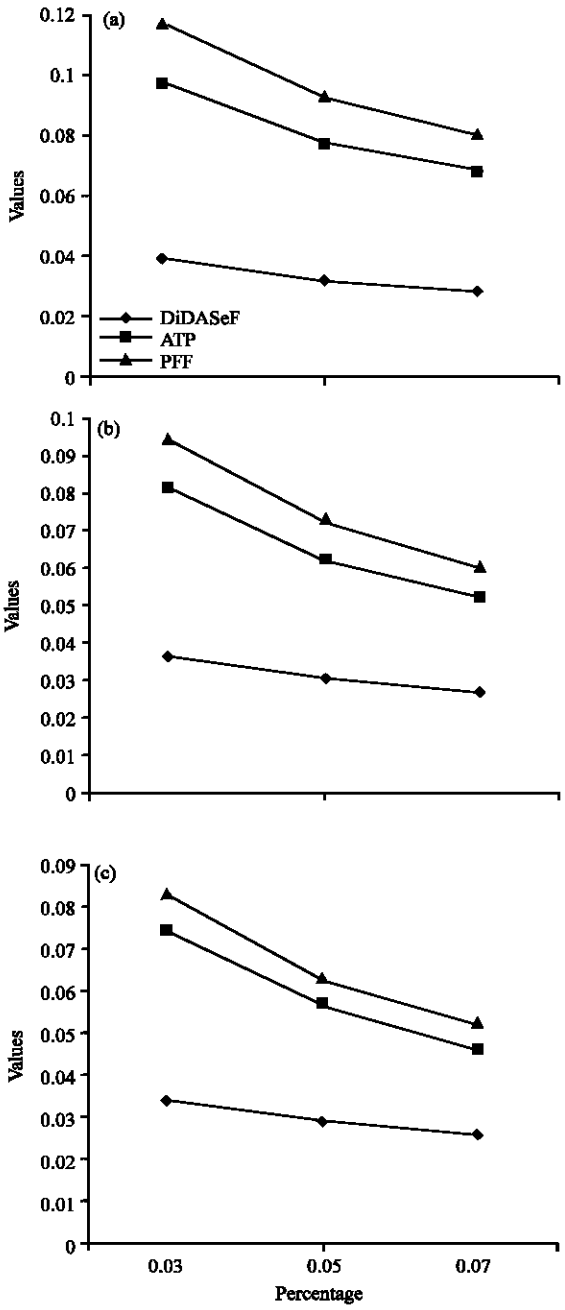

Fig. 5: Energy consumption (loules) at each sensor node: a) p = 20; b) p = 50 and c) p = 100

minimizes when ρ and δ(ε) maximizes because of using efficient method for data reduction in the first stage whilst the selective forwarding in the second stage prevent transmitting the replicated data sets to the cluster head.

Energy consumption: The energy consumption is another performance factor for evaluating our protocol in comparison with other existing methods (ATP and PFF). The energy consumption minimized when the transmitted data to the cluster head minimized. DiDASeF protocol decreases the consumed energy while maintaining the integrity of information by applying two energy efficient approaches data aggregation and selective forwarding. Figure 5 displays the energy consumption for DiDASeF,

ATP and PFF for various $\delta(\varepsilon)$ and ρ values. The conducted simulation results explain that DiDASeF is better than both ATP and PFF from the energy consumption point of view. DiDASeF saves more energy when either p or $\delta(\varepsilon)$ increase.

CONCLUSION

The increased use of the sensor networks in several applications led to provide a huge amount of data which are transmitted across the network. This is greatly contributed in decreasing the network lifetime due to the high communication cost. Therefore, the energy efficient data aggregation and communication approaches are very necessary for eliminating the replicated data in the network. In this study, we propose a protocol named DiDASeF (Distributed Data Aggregation and Selective Forwarding) to extend the lifetime of the WSNs. This protocol is periodic and it works into two phases: energy efficient data aggregation using an Adaptive Piecewise Constant Approximation (APCA) and selective forwarding by using Dynamic Time Warping (DTW) distance measure to allow for each sensor node to check the similarity between the successive sensed data sets to prevent sending the current set which is similar to the previous sensed set, thus improving the network lifetime. The simulation results that based on real data of the sensor network using OMNet++ network simulator show that DiDASeF protocol outperforms the ATP and PFF protocols in terms of data reduction percentage, transmitted data sets percentage to the cluster head, energy consumption and data accuracy.

RECOMMENDATIONS

In future, we plan to apply the data aggregation into two levels: sensor node level and aggregator node level (cluster heads). The first level is responsible for temporal correlation among the data inside the sensor node whilst the second level deals with data correlation among neighboring nodes.

REFERENCES

Anastasi, G., M. Conti, M. di Francesco and A. Passarella, 2009. Energy conservation in wireless sensor networks: A survey. Ad Hoc Networks, 7: 537-568.

Bahi, J.M., A. Makhoul and M. Medlej, 2014. A two tiers data aggregation scheme for periodic sensor networks. Ad Hoc Sens. Wirel. Netw., 21: 77-100.

Cassisi, C., P. Montalto, M. Aliotta, A. Cannata and A. Pulvirenti, 2012. Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining. INTECH, Rijeka, Croatia, ISBN:9789535107484,.

Chong, L., K. Wu and J. Pei, 2007. An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. IEEE Trans. Parallel Distrib. Syst., 18: 1010-1023.

Dalbro, M., E. Eikeland, A.J.I. Veld, S. Gjessing and T.S. Lande *et al*., 2008. Wireless sensor networks for off-shore oil and gas installations. Proceedings of the 2nd International Conference on Sensor Technologies and Applications SENSORCOMM'08, August 25-31, 2008, IEEE, Cap Esterel, France, ISBN:978-0-7695-3330-8, pp: 258-263.

Gedik, B., L. Liu and S.Y. Philip, 2007. ASAP: An adaptive sampling approach to data collection in sensor networks. IEEE. Trans. Parallel Distrib. Syst.,1 18: 1766-1783.

Harb, H., A. Makhoul, A. Jaber, R. Tawil and O. Bazzi, 2016. Adaptive data collection approach based on sets similarity function for saving energy in periodic sensor networks. Intl. J. Inf. Technol. Manage., 15: 346-363.

Harb, H., A. Makhoul, R. Couturier and M. Medlej, 2015. ATP: An aggregation and transmission protocol for conserving energy in periodic sensor networks. Proceedings of the IEEE 24th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), June 15-17, 2015, IEEE, Larnaca, Cyprus, ISBN:978-1-4673-7692-1, pp: 134-139.

Harb, H., A. Makhoul, R. Tawil and A. Jaber, 2014. Energy-efficient data aggregation and transfer in periodic sensor networks. IET. Wirel. Sens. Syst., 4: 149-158.

Heinzelman, W.R., A. Chandrakasan and H. Balakrishnan, 2000. Energy-efficient communication protocol for wireless microsensor networks. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, January 4-7, 2000, Maui, HI., USA., pp: 1-10.

Idrees, A.K., K. Deschinkel, M. Salomon and R. Couturier, 2015. Distributed lifetime coverage optimization protocol in wireless sensor networks. J. Supercomput., 71: 4578-4593.

Idrees, A.K., K. Deschinkel, M. Salomon and R. Couturier, 2016. Perimeter-based coverage optimization to improve lifetime in wireless sensor networks. Eng. Optim., 48: 1951-1972.

Mainwaring, A., D. Culler, J. Polastre, R. Szewczyk and J. Anderson, 2002. Wireless sensor networks for habitat monitoring. Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, September 28, 2002, ACM, Atlanta, Georgia, ISBN:1-58113-589-0, pp: 88-97.

Povedano, S.P., R.A. Valles and J.C. Sueiro, 2014. Selective forwarding for energy-efficient target tracking in sensor networks. Signal Process., 94: 557-569.

Ren, F., J. Zhang, Y. Wu, T. He and C. Chen *et al*., 2013. Attribute-aware data aggregation using potential-based dynamic routing in wireless sensor networks. IEEE. Trans. Parallel Distrib. Syst., 24: 881-892.

Sharaf, M.A., J. Beaver, A. Labrinidis and P.K. Chrysanthis, 2003. TiNA: A scheme for temporal coherency-aware in-network aggregation. Proceedings of the 3rd ACM International Workshop on Data Engineering for Wireless and Mobile Access, September 19, 2003, ACM, San Diego, California, ISBN:1-58113-767-2, pp: 69-76.

Tran, K.T.M. and S.H. Oh, 2014. Uwsns: A round-based clustering scheme for data redundancy resolve. Intl. J. Distrib. Sens. Netw., 2014: 1-6.

Villas, L.A., A. Boukerche, H.A.D. Oliveira, R.B.D. Araujo and A.A. Loureiro, 2014. A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks. Ad Hoc Netw., 12: 69-85.

Wang, Y., P. Wang, J. Pei, W. Wang and S. Huang, 2013. A data-adaptive and dynamic segmentation index for whole matching on time series. Proc. VLDB. Endowment, 6: 793-804.

Xu, X., X.Y. Li, P.J. Wan and S. Tang, 2012. Efficient scheduling for periodic aggregation queries in multihop sensor networks. IEEE. ACM. Trans. Networking, 20: 690-698.

Yahmed, Y.B., A.A. Bakar, A.R. Hamdan, A. Ahmed and S.M.S. Abdullah, 2015. Adaptive sliding window algorithm for weather data segmentation. J. Theor. Appl. Inf. Technol., 80: 322-333.

Zhai, X. and T. Vladimirova, 2015. Data aggregation in wireless sensor networks for lunar exploration. Proceedings of the 6th International Conference on Emerging Security Technologies (EST), September 3-5, 2015, IEEE, Braunschweig, Germany, ISBN:978-1-4673-9799-5, pp: 30-37.