Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
Published Online First: January 2022 **2022, 19(4): 875-886** **E-ISSN: 2411-7986**

DOI: http://dx.doi.org/10.21123/bsj.2022.19.4.0875

Perceptually Important Points-Based Data Aggregation Method for Wireless Sensor Networks

Iman Dakhil Idan Saeedi ***Ali Kadhum M. Al-Qurabat*** *

Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq.
*Corresponding author: alik.m.alqurabat@uobabylon.edu.iq
E-mail address: iman.idan@student.uobabylon.edu.iq

Received 18/2/2021, Accepted 24/5/2021, Published Online First 20/1/2022, Published 1/8/2022

Abstract:

The transmitting and receiving of data consume the most resources in Wireless Sensor Networks (WSNs). The energy supplied by the battery is the most important resource impacting WSN's lifespan in the sensor node. Therefore, because sensor nodes run from their limited battery, energy-saving is necessary. Data aggregation can be defined as a procedure applied for the elimination of redundant transmissions, and it provides fused information to the base stations, which in turn improves the energy effectiveness and increases the lifespan of energy-constrained WSNs. In this paper, a Perceptually Important Points Based Data Aggregation (PIP-DA) method for Wireless Sensor Networks is suggested to reduce redundant data before sending them to the sink. By utilizing Intel Berkeley Research Lab (IBRL) dataset, the efficiency of the proposed method was measured. The experimental findings illustrate the benefits of the proposed method as it reduces the overhead on the sensor node level up to 1.25% in remaining data and reduces the energy consumption up to 93% compared to prefix frequency filtering (PFF) and ATP protocols.

Key words: Data Aggregation, Energy-Saving, Perceptually Important Points (PIP), Wireless Sensor Network.

Introduction:

Wireless Sensor Networks (WSNs) can be utilized in a wide range of applications, including military surveillance and environmental and facility monitoring[1,2]. A WSN is usually made up of multiple sensor nodes with communication capabilities, as well as the ability to connect with any external sinks or base stations[3,4]. These sensors are either dispersed at random around a rugged landscape (such as a battlefield) or are strategically placed. Different communication networks, such as single and multi-hop networks, or a hierarchically organized system with a number of clusters and cluster heads, are formed by the synchronization of these sensors[5]. The sensor senses, processes, and transmits data to the base stations on a regular basis. The frequency at which data is reported, as well as the number of sensors involved in the operation, are both dictated by the application[6].

Data gathering is the act of routinely extracting data from a variety of sensors and sending it to the base station for processing. Giving the sensor node's energy constraints, direct data transfer to the base station by all sensors will be inefficient[7]. This is attributed to the high overlap of data from neighboring sensors, which results in redundancy. Furthermore, base stations are incapable of processing the massive quantities of data provided by a wider sensor network[8].

As a result, such networks are expected to merge data and generate significant information at sensors or intermediate nodes, as this can help to reduce packet transfers to the base station, saving energy and bandwidth. Data aggregating techniques, which typically require the fusion of data collected from multiple sensors at intermediate nodes and routing aggregated data to base stations, may be used to do this[9].

The major contributions in this paper concentrate on the design and application of a strategy for energy-efficient data aggregation to extend the lifespan of WSNs. The contributions that this paper provides are as follows:

1. At the sensor node level, a method of data aggregation based on the perceptually important points is suggested for reducing the amount of data readings transmitted, reducing

the energy consumed and thereby extending the lifespan of the network whereas preserving the accuracy of the data readings obtained at the base station.

2. The evaluation of the suggested approach is carried out with the use of comprehensive simulation experiments provided by the simulator of the OMNeT++ network. The efficiency of the proposed technique is evaluated with two related works: the PFF protocol proposed in[10] and the ATP protocol proposed in[11].

The reaming portion of the research paper is systematized as follows: Section II surveys existing works on data aggregation methods. Section III focuses on the proposed method that aggregates data based on the perceptually important points. In Section IV & V results and conclusions are drawn.

Related Works:

For WSN, a large number of studies have addressed the reduction of data transmissions through removing data redundancy (i.e., using data aggregation). The primary target of this review is to thoroughly examine the published works of literature on extending the lifespan of WSN[12-33]. There are several methods and principles dedicated to save energy and expanding WSN's lifespan, concentrating mostly on reducing data transfer, such as predictive monitoring, routing, aggregation, elimination, prediction, adaptive sampling, clustering and data compression. Data Aggregation before transmitting it is thus a key strategy in terms of energy efficiency.

A description of the related works on the techniques of data aggregation in WSN is shown in Table 1.

The Proposed Data Aggregation Method:

This section will include a comprehensive explanation of the proposed process, which is a technique of data aggregation based on the perceptually important points for decreasing the amount of transmitted data readings, reducing the energy consumed and thereby extending the network lifetime while retaining the precision of the data readings obtained at the base station.

WSN Topology:

The PIP-based data aggregation technique was developed based on cluster topology. The WSN-based clustering is shown in Fig. 1. In the proposal, the formation of the cluster topology is out of scope, it is assumed that there is already a topology and deliberately skipping to discuss the formation of the topology. The proposed technique can be applied to these clusters produced by any clustering protocol. Focusing basically on designing an energy-efficient data aggregation method. More precisely, the objective of this technique is to reduce the sensed data at the sensor node level to prolong the WSN's lifetime.

Table 1. A summary of data aggregating mechanisms and their main characteristics.

Year	Network Type	Node Type	Aggregator	Algorithm	Application type	Remarks
(12) 2014	Tree	Homogenous	Fixed	Distributed	Monitoring Applications	GSTEB, the sink broadcasts root node collection based on itself and the info nodes of the neighbor are connected to the parent nodes.
(13) 2015	Cluster	Homogenous	Mobile	Distributed	Health and underwater monitoring	It has a mobile connectivity and rendezvous node that serves as a mobile connection store.
(14) 2015	Cluster, Flat	Heterogeneous	Fixed	Centralized, Local	Monitoring Applications	To achieve efficient data aggregation, tradeoffs between storage efficiency and energy dissipation must be made. The algorithm is implemented in two ways: Distributed Energy Allocation (DBA) and Centralized Energy Allocation (CEA). It makes use of a GA-based methodology as well as Gibbs Sampling Theory.
(15) 2016	Cluster, Tree	Homogenous	Fixed	Distributed	Emergency/ Critical operation	Using a priority-based dynamic data aggregation scheme, a three-layer big data aggregation architecture has been created (PDDA).

(16) 2016	Flat	Homogeneous	Fixed	Centralized	Environment monitoring	The data is first filtered using the fitting algorithm, which has two thresholds; if the data's usual value is less than one of the thresholds, the data is not sent to the aggregator for aggregation.
(17) 2017	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	Collects sensor readings on a regular basis to extend the life of a Periodic Sensor Network (PSN).
(8) 2018	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	Delete the redundant data gathered and adjust the sampling rate to match the monitored environmental conditions.
(18) 2018	Cluster	Homogeneous	Fixed	Distributed	Precision Agriculture applications	Redundant data transfer is minimized during sequential iterations by using differential data from the sensor.
(9) 2018	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	An Adaptive Piece-wise Constant Approximating approach is used to combine and use data dimensionality.
(19) 2018	Flat	Homogeneous	Mobile	Distributed	Monitoring applications	For efficient route selection of the mobile node, Adaptive Particle Swarm Optimization is used.
(20) 2019	Flat	Homogeneous	Fixed	Distributed	Monitoring applications	To define and eradicate data duplication, the aggregator employs a linear classifier based on SVM.
(21) 2019	Cluster	Homogeneous	Mobile	Distributed	Disaster-prone zone monitoring	Cluster creation based on the Neural network and optimization based on Ant Colony.
(22) 2019	Tree	Homogeneous	Fixed	Distributed	Wireless water meter application	In addition to the meta-heuristic algorithm for minimizing energy and decreasing latency, multi-channel TDMA scheduling algorithms are used to reduce collisions.
(5) 2019	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	For energy-saving the authors suggested data aggregation in WSNs, utilizing Integrated Divide and Conquer with improved K-means technique.
(23) 2019	Cluster	Homogeneous	Fixed	In-Network	Environment monitoring	In wireless Sensor Networks, local aggregation algorithms use temporally correlating functionality to eliminate data redundancy and local outlier data, thus improving data consistency and transmission ratios.
(24) 2020	Cluster	Homogeneous	Fixed	In-Network	Environment monitoring	The homogeneous data is divided into clusters, and the in-network data is reduced by selecting the core values for each cluster.
(25) 2020	Cluster	Homogeneous/ heterogeneous	Fixed	Distributed	Environment monitoring	Provide a modern form of data aggregation focused effectively on the open-pit mining concept.
(26) 2020	Cluster	Homogeneous	Fixed	Distributed	Disaster Management and rescue operation	For real-time data analysis, a cluster-based systematic data aggregation paradigm (CSDAM) is proposed.
(1) 2020	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	To get rid of the redundant data readings, the SAX symbolic algorithm and adaptive piecewise constant approximation (APCA) were used as a data aggregation technique.
(27) 2021	Cluster	Homogeneous	Fixed	Distributed	Environment monitoring	A clustering-based Dynamic Time Warping (DTW) algorithm is used in fog computing, and an optimized grouping and basic encoding algorithm is used in sensor nodes.

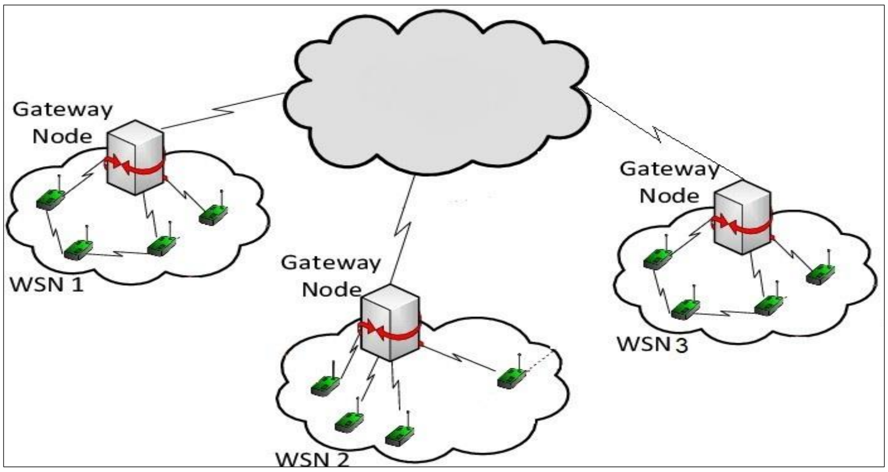

Figure 1. WSN-based clustering topology for IoT network.

In cluster-based architecture, every cluster has several cluster members (sensor nodes) and one common cluster head (CH). Usually, when the sensor nodes are scattered in close regions, they can generate the same readings or similar readings, and therefore the temporal and spatial correlation can be exploited. The data set of the sensor node is sent to the respective CH that belongs to it after the end of each period.

In the proposed method, the following assumptions have been considered regarding WSN topology:

- All sensors are homogeneous, predetermined deployed within the communication range of the base station node.
- For energy consumption, each node is presumed to be using the same model of radio.
- Each node is presumed to use a periodic data collection mode, where the data collected is processed and sent to the corresponding CH regularly by each node.
- Data transfer from the sensor nodes to the relevant CH relies on single-hop communication.
- Environmental conditions or events such as pressure and temperature are monitored by a sensor node.
- It is possible to partition a cluster-based network into disjoint clusters. There is one cluster head (CH) and several sensor nodes (SNs) in each cluster. Each CH gathers data from its SNs and transmits the processed data to the base station.

Data Collection:

The main objective of WSN is to make human life easier and simpler. The implementation of WSN is often concerned with data collection and communication of information. In WSN context, the data is often collected from sensors. Based on application requirements, in WSN, the collection of data may be event-driven (like forest fire, oil and gas leaks detection) or time-driven (like habitat monitoring, logging temperature and humidity in the plants for precision agriculture). The time-driven data collection model, called Periodic, is taken into account in this article.

Each sensor node i captures a data readings vector R for each cycle and then transmits it to the CH in the periodic applications as follows: $R_i = [r_1, r_2, \dots, r_{\tau-1}, r_\tau]$ where τ reflects the total number of data readings obtained in the period ρ. Figure 2 displays a periodic data collection example in which every sensor node capture one reading of data every 10 minutes, e.g. s = 10 minutes, and transmit the set of collected data that include 6 reads, e.g. τ = 6, to CH at end of every hour.

Often, data readings obtained from the sensor are redundant in any cycle, i.e., in R_i, depending on how the conditions monitored differ. In order to minimize the amount of data readings transmitted and to preserve the energy of the sensor, the search for data redundancy in each sensor is therefore important. Hence, our objective is to reduce the size of R_i by aggregating it using the perceptually important points (PIP) segmentation method.

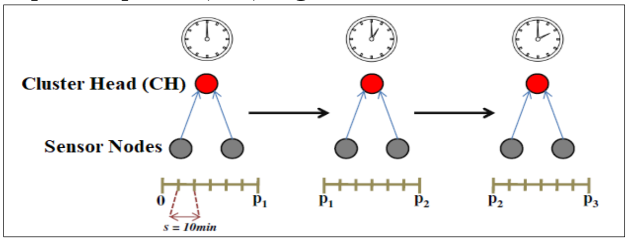

Figure 2. Illustrative example of Periodic data collection.

Perceptually Important Points Based Data Aggregation (PIP-DA) Method:

Using PIPs is a promising approach to manipulate salient points from a time series. In the data mining context, PIPs were used mostly for dimensional reduction (time series representation) purposes, as a hierarchical method for time series segmentation[34] and for clustering purposes[35].

Time series are formed from a series of readings of sensor data $R_i = [r_1, r_2, \dots, r_{\tau-1}, r_\tau]$ and the importance of each reading r_i has varying degrees of effect on the time series of the sensor's motion form. That is, each data reading will have its own meaning for the time series, the total motion form of the time series can be calculated by a data reading, while another either has no effect on the time series or can even be discarded. PIP technology aims to find the reading that has a crucial effect on the overall time series motion form[36].

According to the proposed data aggregation method, a collected sensor data reading $R_i = [r_1, r_2, \dots, r_{\tau-1}, r_\tau]$ can be expressed by a PIP series $\mathcal{P} = [\wp_1, \wp_2, \dots, \wp_n]$, where $n \ll \tau$ as follows:

- The PIP-DA begins by identifying as the first two PIPs $\wp_1$ and $\wp_2$ the first r_1 and last r_τ readings of the original sensor data readings R.
- Next, it computes the distance between all residual readings of sensor data and the two initial PIPs.
- Subsequently, the PIP-DA chooses the sensor data reading r_i with the maximum distance as the third PIP $\wp_3$.
- The PIP-DA selects as the fourth PIP $\wp_4$ the sensor data reading that significantly increases its distance from its neighbouring PIPs (that are either the first $\wp_1$ and the third $\wp_3$, or the PIP identified by the third $\wp_3$ and the second $\wp_2$).
- The PIP-DA will stop when the number of PIPs specified by the user is determined.

The Measures of Distance for PIP Detection:

In the PIP-DA technique, three metrics can be utilized for distance, namely vertical distance (VD), perpendicular distance (PD) and Euclidean distance (ED) as shown in Fig. 3.

This paper has used the Euclidean distance (ED) as a measure for the PIP detection. Let $R = [r_1, r_2, \dots, r_{\tau-1}, r_\tau]$ be the sensor data readings of length τ, and two adjacent PIP $\wp_u = r_u$ and $\wp_v = r_v$. The Euclidean distance of each of the intermediate data readings $\wp_i = r_i$, for $i \in \{u + 1, \dots, v - 1\}$ from the two PIPs is defined as in Eq.1.

$$ED(\wp_i, \wp_u, \wp_v) = \sqrt{(\wp_u - \wp_i)^2} + \sqrt{(\wp_v - \wp_i)^2} \qquad 1$$

The new point of PIP is the one which maximizes the distance, $\wp = \text{argmax}_i(ED(\wp_i, \wp_u, \wp_v))$, where "argmax" refers to the maximum argument.

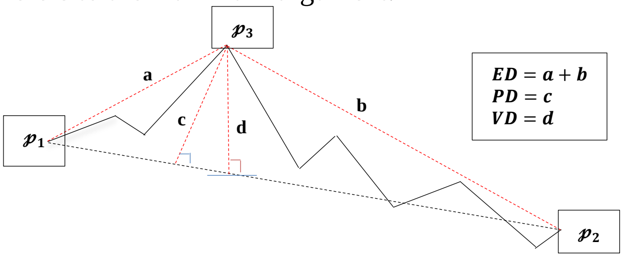

Figure 3. Detection of the 3rd PIP with three different measures of distance VD, PD and ED.

Application Criticality:

Using different types of sensor instruments, such as noise detection, concentration of chemicals, strain, displacement and temperature, the WSN may be used to track disasters. The effect on individuals and on the ecosystem of disasters is not the same. Therefore, if the disaster risk level is high, the sensor node would send more data readings than if the disaster risk level is low which will include high quality collected data readings to simplify both the analysis and to help understand the tracked catastrophe.

The number of PIP in proposed method depends on the application criticality (i.e., the disaster risk level). There is a proportional relation between PIPs number and the risk level; therefore, when the risk level is high, the PIP-DA will push the sensor node to send as maximum as possible data readings and vice versa. In general, this would save energy because the sensor node has the option to adjust the sending rate based on the needs of the application in WSNs. The criticality of the application is represented in the PIP-DA method as a minimum number of transmitted data readings for a sensor node, $\mathcal{A}_{\mathcal{CR}}$, over a period. $\mathcal{A}_{\mathcal{CR}}$ takes values from 1 to 100, indicating either a low or high criticality level, respectively. So, the number of PIP is calculated as in Eq2.

$$PIP_{No} = (No. of\ collected\ Sensor\ data\ readings \times \mathcal{A}_{\mathcal{CR}})/100 \qquad 2$$

All the sensor nodes initialize their $\mathcal{A}_{\mathcal{CR}}$ based on the specifications of the application and before deployment. During the lifespan of the network, it is also possible to dynamically adjust $\mathcal{A}_{\mathcal{CR}}$ for all sensors or only for a specified subgroup of sensors if certain forms of management and control schemes are usable.

Algorithm 1 illustrates the PIP-DA method. The main purpose of this algorithm is to reduce the number of transmitted data readings to the CH. The

PIP is used to segment the sensor data readings dynamically.

Algorithm 1: PIP-Data Aggregation Method

Input: Sensor data readings $\boldsymbol{R} = [r_1, r_2, \ldots, r_{\tau-1}, r_\tau]$; Application Criticality: $\boldsymbol{\mathcal{A}_{\mathcal{CR}}}$ $\boldsymbol{PIP}$ series $\boldsymbol{\mathcal{P}} = [\wp_1, \wp_2, \ldots, \wp_n]$

Output :

$\boldsymbol{for\ i} \leftarrow 1\ \boldsymbol{to}\ \tau\ \boldsymbol{Do}$
  $R \leftarrow Captures\ sensor\ data\ reading$
$\boldsymbol{end\ for}$
$\boldsymbol{n = (R \times \mathcal{A}_{\mathcal{CR}})/100}$
$\boldsymbol{Set}\ \wp_1 = r_1;\ \wp_2 = r_\tau$
$\boldsymbol{PIP_{No} = 2}$
$\boldsymbol{while}\ (PIP_{No} \le n)\ \boldsymbol{do}$
  $Calculate\ the\ Euclidean\ Distance\ ED_i\ of\ residual\ readings\ r_i\ in\ R$
  $Select\ r_i\ with\ maximum\ ED_i\ as\ next\ PIP\ \wp_i$
  $Append\ PIP \leftarrow \wp_i$
  $PIP_{No}++$
$\boldsymbol{end\ while}$
$Arrange\ \boldsymbol{PIP}\ according\ to\ the\ index\ of\ se$
$Encode\ \boldsymbol{PIP}$
$return\ \boldsymbol{PIP}$

Finally, in order to further reduce the size of the resulting PIP series before it is sent to the CH, the proposed method will encode this PIP series; where, as shown in Fig. 4, each $\wp_i$ in the PIP series is encoded using two bytes. From Fig. 4, every $\wp_i$ is encoded in a 16-bit representation in the PIP series. For negative numbers, the sign bit takes 1 and positive numbers 0. The integer part of the $\wp_i$ was expressed in the 8 bits that followed. The remaining 7 bits constitute the fraction part of the $\wp_i$.

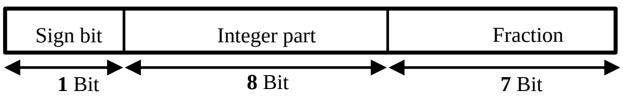

Figure 4. Binary encoded PIP series.

Simulation Experiments and Results:

This section shows the performance evaluation and simulation results as graphs and discussion for the proposed technique outlined in Section 3. The goal is twofold: first, evaluating the performance of technique with different performance metrics via real sensor data. In Table 2, the efficiency of PIP-DA is measured using the following parameters. Second, comparing the technique proposed with recent existing protocols belongs to the same field.

Table 2. The settings of parameters.

Parameters	Value
WSN size	47 sensors
Sensor data readings $\boldsymbol{R}$	20, 50 and 100 readings
Application Criticality $\boldsymbol{\mathcal{A}_{\mathcal{CR}}}$	5%, 10%, 25%, 50% and 75%
Similarity Threshold $\boldsymbol{\delta}$	0.03, 0.05 and 0.07
E_{elec}	50 nJ/bit
ε_{amp}	100 $pJ/bit/m^2$

Simulation Environment:

To evaluate proposed techniques, extensive simulation experiments are carried out using the OMNeT++ simulator and dependent on actual data from sensors. A network of N sensors and a single-hop topology was considered installed in the laboratory during these simulations. The middle of the laboratory comprises a single CH node. This installation is shown in Fig. 5.

Periodically, sensors measure the local measurements at a set frequency (e.g., temperature). Proposed technique is disseminated in every sensor node, which is dependent on using the Intel Berkeley Research laboratory dataset[37]. These sensed weather data (like light, humidity and temperature) are collected periodically every 31 seconds. The sensor nodes using a log file in our simulations that contains 2.3 million readings previously obtained by 54 Mica2Dot sensor nodes in the lab as shown in Fig. 5. This paper only uses one measure of measurements of sensor nodes: temperature. Each sensor node shown with a yellow sign in Fig. 5 is not included in our experiments because its data may be incomplete or truncated. Then the temperature readings for 47 sensor nodes are collected and processed. The findings are 47 sensor nodes averaging.

Baghdad Science Journal
2022, 19(4): 875-886
P-ISSN: 2078-8665
E-ISSN: 2411-7986

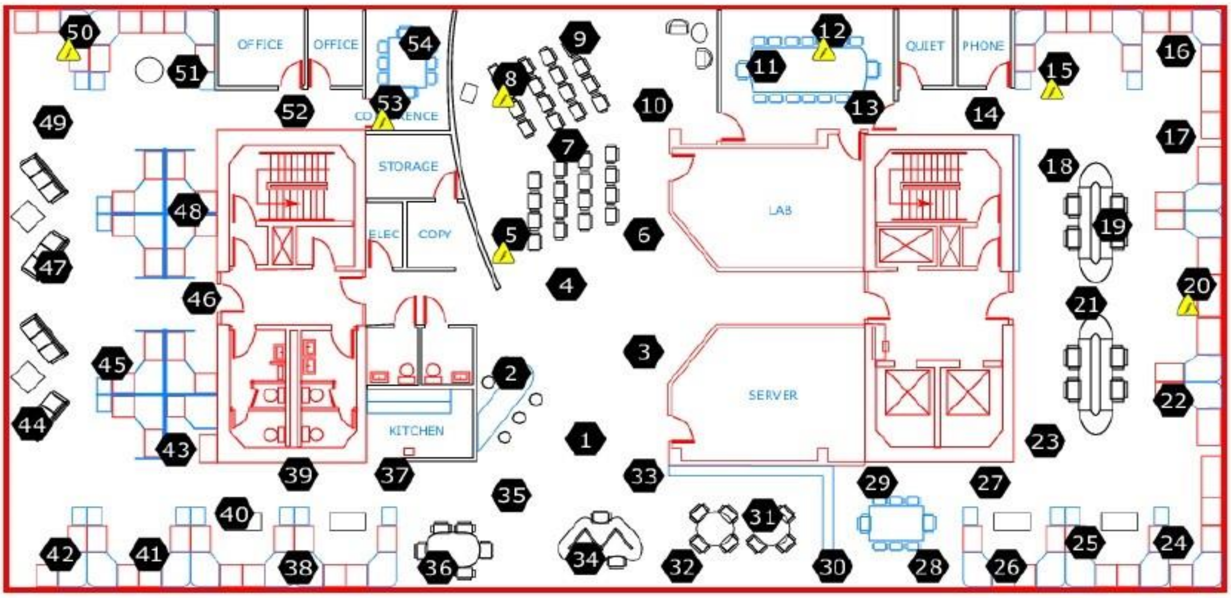

Figure 5. Deployment of Sensors in the Intel Berkeley Lab.

The Residual Data after Aggregation:

Within this experiment, the aim is to demonstrate how the proposed technique can be used by the sensor nodes to aggregate the collected data readings (i.e., remove redundant data readings at each period). Figure 6 indicates the proportion of residual data readings that will be remained once the redundancy has been removed by applying the PIP method. It may be easily seen that with the various parameters, the proposed technique has the ability to adjust the sending rate based on the application criticality.

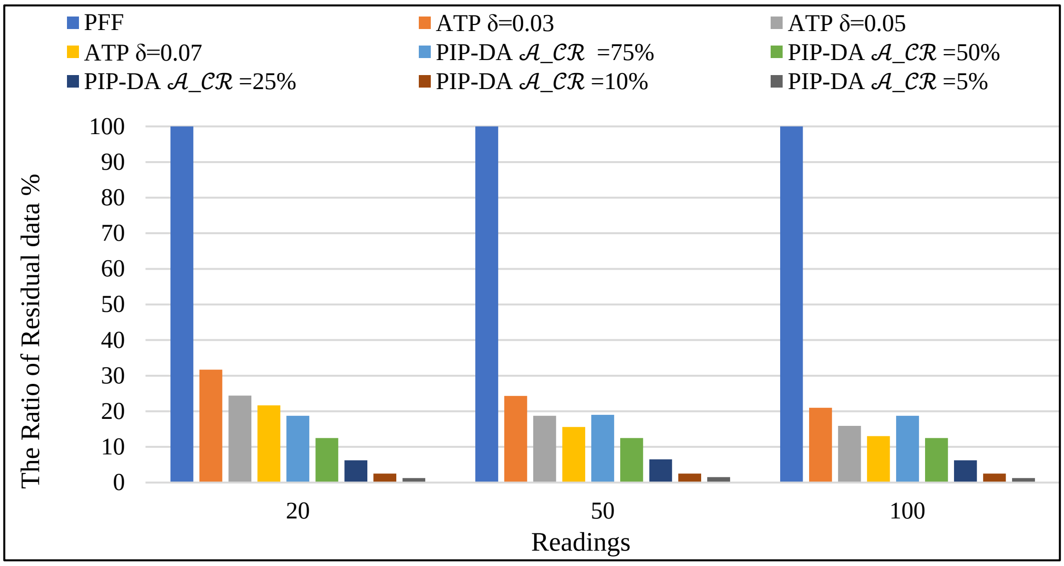

Figure 6. The Ratio of Residual Data after Aggregation.

Based on the findings obtained, it can infer the following things:

- The outcome of the aggregation at the proposed strategy PIP-DA depends on the application criticality level $\mathcal{A}_{\mathcal{CR}}$ selected. As the greater the risk level of the application, the greater the amount of data residual PIP and vice versa.
- ATP is found to keep smaller amounts of data if the amount of data collected R or the similarity threshold δ increases.
- While the PFF keeps all the collected data 100%.

The Ratio of Transmitted Data Sets:

In this experiment, based on the suggested PIP-DA system, each sensor decreases the number of data sets transmitted to its respective CH. Figure 7 indicates the ratio of data sets transmitted utilizing PIP-DA, PFF and ATP protocols by a sensor node. The PIP-DA helps each sensor node, depending on

Baghdad Science Journal
2022, 19(4): 875-886
P-ISSN: 2078-8665
E-ISSN: 2411-7986

the degree of application criticality, to change its transmission rate.

Several assumptions can be rendered as follows, based on the findings in Fig. 7:

- As δ and $\mathcal{A}_{\mathcal{CR}}$ are raised in the ATP and PIP-DA methods, the sensor node sends further sets.
- Our proposed method sends fewer data sets than ATP and PFF protocols in all the cases when varying the application criticality level $\mathcal{A}_{\mathcal{CR}}$ between 5% to 75%.
- While the PFF sends all the collected data 100%.

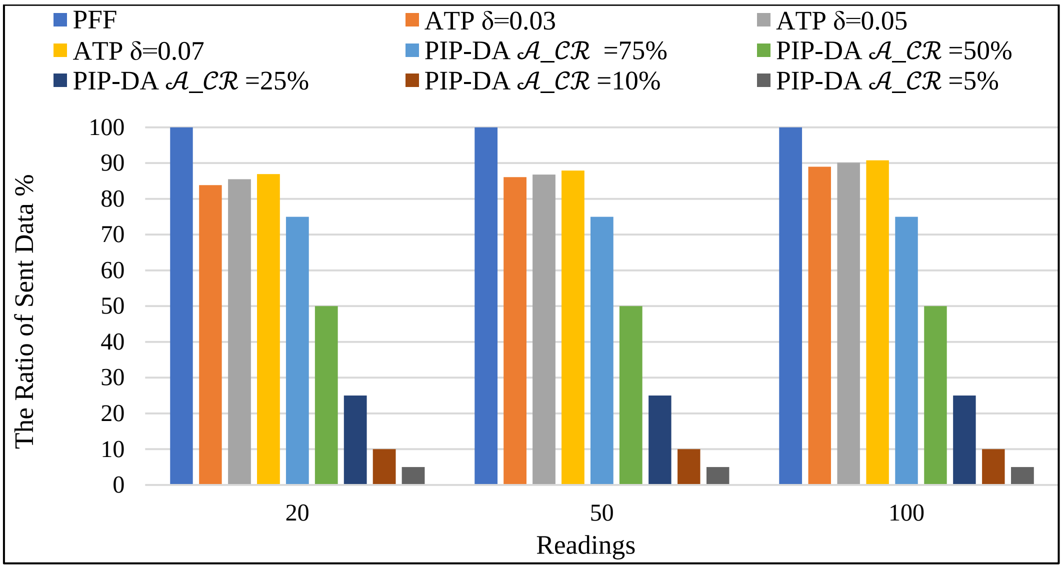

Figure 7. The Ratio of Transmitted Data Sets.

The Analysis of Energy Consumption:

Our aim in this experiment is to research the cost of energy at the sensor node level. At the sensor node, the energy consumed represents the energy consumed in sending data to the CH. The same model for energy indicated in[27,28,29,30] is used. The energy consumption of the transmission demands extra power to amplify the signal due to its distance from the endpoint. Thus, the radio consumes energy as defined in Eq. 3 to send a $\mathcal{K}$-bit message to distance $\mathcal{D}$, where E_{elec} is the power required by radio electronics and is equivalent to $50\,nJ/bit$, ε_{amp} is the power required by the amplifier and is equivalent to $100\,pJ/bit/m^2$.

$$E_{TX} = E_{elec} \times \mathcal{K} + \varepsilon_{amp} \times \mathcal{K} \times \mathcal{D}^2 \qquad 3$$

Figure 8 shows a comparison between our technique PIP-DA, ATP and PFF in terms of the amount of energy consumed using different parameters. The findings obtained indicate our technique's dominance over ATP and PFF by reducing energy consumption.

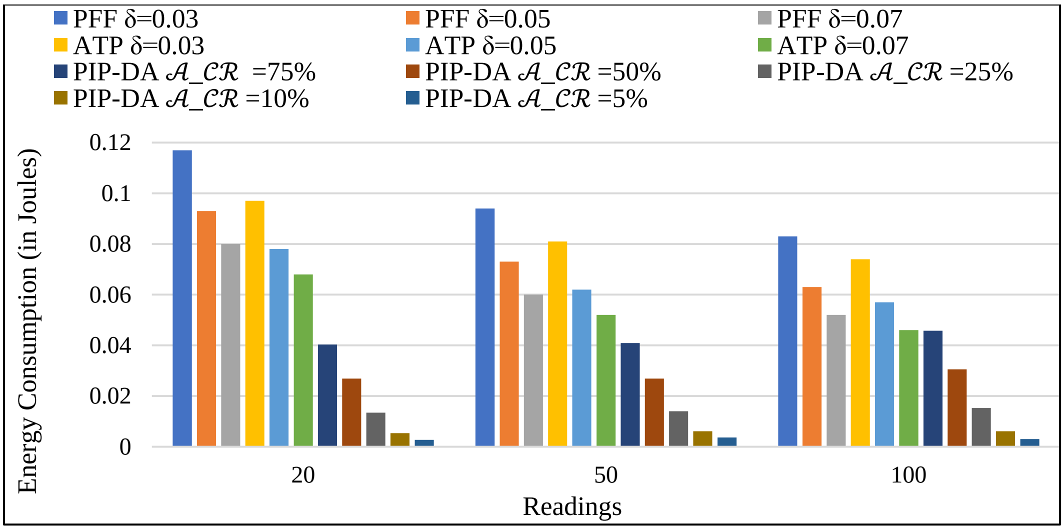

Figure 8. The Energy Consumption.

Based on the findings obtained, it can infer the following points:

- In the proposed technique, the reduction of redundant sets has a very large impact on reducing energy consumption by reducing the operation of the radio unit (i.e., transmission and reception operations).
- The application criticality level $\mathcal{A}_{\mathcal{CR}}$ plays an influential role in energy consumption, as increasing the $\mathcal{A}_{\mathcal{CR}}$ leads to an increase in the energy consumption; the reason is due to sending more packets.
- Also, the data size captured R and similarity threshold δ can affect the energy consumption where increasing R and δ will reduce the energy consumption as the case in ATP and PFF.

Data Accuracy:

A significant problem for the WSN is to delete redundant data without compromising accuracy. The Accuracy of data reflects the "loss rate" of readings captured by sensor nodes while the CH does not receive it. Figure 9 indicates the proportion of accuracy (i.e., data loss rate) which will not be delivered to the CH once the data sets have been aggregated. It may be easily seen that, with the various parameters, there is a trade-off between data accuracy, the amount of data transmitted (see Fig. 7) and energy consumption (see Fig. 8). To have high accuracy, more data must be sent and thus more energy spent. Some applications do not need high accuracy in data, such as monitoring the environment, and thus fewer data can be sent, while military and health applications need high accuracy in data and therefore need to send a larger amount of data for this purpose.

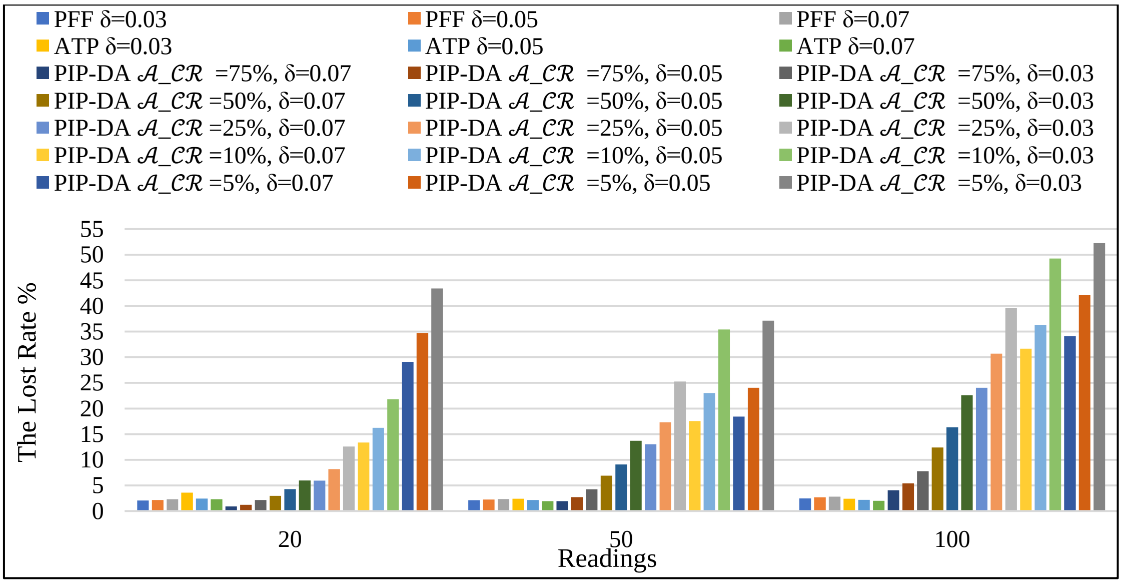

Figure 9. The Accuracy of Data.

Based on the findings obtained, it can infer the following things:

- In the proposed technique PIP-DA the reduction of the data readings has a relation to the accuracy, the loss rate increases when the application criticality level $\mathcal{A}_{\mathcal{CR}}$ decreases.
- The similarity threshold δ plays an influential role in the accuracy of the data, as increasing the similarity threshold δ leads to an increase in the accuracy of the data.
- Also, the data size captured R can affect the accuracy were increasing the volume of data collected R will make the rate of data loss big.
- Note that the accuracy in the ATP and PFF protocols is better than our proposed method PIP-DA for the level of application criticality $\mathcal{A}_{\mathcal{CR}} > 25\%$, but at the expense of more energy expenditure, as shown in Fig. 8.

Conclusion:

Wireless sensor networks are networks with a limited amount of energy. Given that data communication consumes the vast majority of resources, some method of data aggregation is needed. As a result of its ability to remove redundant data transfers within the network, this method has lately attracted a lot of attention. To improve the WSN lifetime, the principal idea is to exploit the advantage of the temporal data correlation between the sensor node data readings to minimize the energy depletion by aggregating sensed data before sending them to the CH. In the sensor node level, a Perceptually Important Points

Based Data Aggregation (PIP-DA) method for WSNs is suggested for reducing the number of transmitted data readings, decreasing the consumed energy, and thus extending the network lifespan whereas preserving the accuracy of the data reading received at the base station. The simulation results that based on real data of the sensor network using OMNeT++ simulator show that the proposed data aggregation approach outperforms some recent existing approaches in terms of several performance metrics like remaining data after aggregation, the ratio of sets sent to the CH, data accuracy at CH and energy consumption.

For future studies, it is recommended to investigate the possibility of implementing another dynamic segmentation algorithm that can be applied at two levels: the sensor node and gateway. In order to forecast the missing data at the CH, also expect to implement a prediction approach and combine it with our work.

Acknowledgments:

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

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Babylon.

Authors' contributions:

Ali K. M. Al-Qurabat and Iman D. I. Saeedi contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

References:

1. Al-Qurabat AKM, Idrees AK. Data gathering and aggregation with selective transmission technique to optimize the lifetime of Internet of Things networks. *Int. J. Commun. Syst.* 2020; 33(11):e4408. DOI: 10.1002/dac.4408
2. Abdulzahra SA, Al-Qurabat AKM, Idrees AK. Data Reduction Based on Compression Technique for Big Data in IoT. In: *2020 International Conference on Emerging Smart Computing and Informatics, ESCI* 2020. 2020 (pp. 103-108). IEEE. DOI: 10.1109/ESCI48226.2020.9167636
3. Al-Qurabat AKM, Abdulzahra SA. An overview of periodic wireless sensor networks to the internet of things. *IOP Conf Ser Mater Sci Eng*. 2020; 928(3):032055. DOI: 10.1088/1757-899X/928/3/032055
4. Idan Saeedi ID, Al-Qurabat AKM. A systematic review of data aggregation techniques in wireless sensor networks. *J Phys Conf Ser*. 2021; 1818(1):012194. DOI: 10.1088/1742-6596/1818/1/012194
5. Idrees AK, Al-Qurabat AKM, Abou Jaoude C, Al-Yaseen WL. Integrated divide and conquer with enhanced k-means technique for energy-saving data aggregation in wireless sensor networks. In: *2019 15th International Wireless Communications & Mobile Computing Conference, IWCMC 2019*. 2019 (pp. 973-978). IEEE. DOI: 10.1109/IWCMC.2019.8766784
6. Al-Qurabat AK, Idrees AK. Two level data aggregation protocol for prolonging lifetime of periodic sensor networks. *Wirel. netw*. 2019; 25(6):3623-41. DOI: 10.1007/s11276-019-01957-0
7. Idrees AK, Al-Qurabat AK. Distributed Adaptive Data Collection Protocol for Improving Lifetime in Periodic Sensor Networks. *IAENG Int. J.Com. Sci*. 2017; 44(3).
8. Al-Qurabat AK, Idrees AK. Energy-efficient adaptive distributed data collection method for periodic sensor networks. *Int. J. Internet Technol. Secur. Trans*. 2018; 8(3):297-335. DOI: 10.1504/IJITST.2018.093660
9. Al-Qurabat AK, Idrees AK. Distributed data aggregation and selective forwarding protocol for improving lifetime of wireless sensor networks. *J. Eng. Appl. Sci*. 2018; 13(5):4644-53.
10. Bahi JM, Makhoul A, Medlej M. A two tiers data aggregation scheme for periodic sensor networks. *Adhoc & Sens. Wirel. Ne*. 2014; 21(1).
11. Harb H, Makhoul A, Couturier R, Medlej M. ATP: An aggregation and transmission protocol for conserving energy in periodic sensor networks. In: *2015 IEEE 24th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises 2015*. 2015(pp. 134-139). DOI: 10.1109/WETICE.2015.9
12. Han Z, Wu J, Zhang J, Liu L, Tian K. A general self-organized tree-based energy-balance routing protocol for wireless sensor network. *IEEE Trans. Nucl. Sci*. 2014; 61(2):732-40. DOI: 10.1109/TNS.2014.2309351
13. Mottaghi S, Zahabi MR. Optimizing LEACH clustering algorithm with mobile sink and rendezvous nodes. *AEU- Int. J. Electron. Commun*. 2015; 69(2):507-14. DOI: 10.1016/j.aeue.2014.10.021
14. Xiao S, Li B, Yuan X. Maximizing precision for energy-efficient data aggregation in wireless sensor networks with lossy links. *Ad Hoc Netw*. 2015; 26:103-13. DOI: 10.1016/j.adhoc.2014.11.014
15. Karim L, Al-kahtani MS. Sensor data aggregation in a multi-layer big data framework. In: *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2016*. 2016(pp. 1-7). DOI: 10.1109/IEMCON.2016.7746261

16. Atoui I, Ahmad A, Medlej M, Makhoul A, Tawbe S, Hijazi A. Tree-based data aggregation approach in wireless sensor network using fitting functions. In: *2016 Sixth international conference on digital information processing and communications, ICDIPC 2016*. 2016 (pp. 146-150). DOI: 10.1109/ICDIPC.2016.7470808
17. Al-Qurabat AK, Idrees AK. Adaptive data collection protocol for extending lifetime of periodic sensor networks. *Qalaai Zanist Sci. J*. 2017; 2(2):83-92. DOI: 10.25212/lfu.qzj.2.2.11
18. Khriji S, Raventos GV, Kammoun I, Kanoun O. Redundancy elimination for data aggregation in wireless sensor networks. In: *2018 15th International Multi-Conference on Systems, Signals & Devices, SSD 2018*. 2018(pp. 28-33). DOI: 10.1109/SSD.2018.8570459
19. SreeRanjani NY, Ananth AG, Reddy LS. An energy efficient data gathering scheme in wireless sensor networks using adaptive optimization algorithm. *J. Comput. Theor. Nanosci*. 2018; 15(1112):3456-61. DOI: 10.1166/jctn.2018.7644
20. Yadav S, Yadav RS. Redundancy elimination during data aggregation in wireless sensor networks for IoT systems. In: *Recent trends in communication, computing, and electronics 2019.* 2019 (pp. 195-205). DOI: 10.1007/978-981-13-2685-1_20
21. Sarangi K, Bhattacharya I. A study on data aggregation techniques in wireless sensor network in static and dynamic scenarios. *Innov. Syst. Softw. Eng*. 2019; 15(1):3-16. DOI: 10.1007/s11334-019-00326-6
22. Kumar S, Kim H. Energy efficient scheduling in wireless sensor networks for periodic data gathering. *IEEE access*. 2019; 7:11410-26. DOI: 10.1109/ACCESS.2019.2891944
23. Verma N, Singh D. Local Aggregation Scheme for Data Collection in Periodic Sensor Network. *Int. J. Eng. Adv. Technol*. 2019; 9(2):3583–3588. DOI: 10.35940/ijeat.b2602.129219
24. Alam MK, Aziz AA, Latif SA, Awang A. Error-Aware Data Clustering for In-Network Data Reduction in Wireless Sensor Networks. *Sensors*. 2020; 20(4):1011. DOI: 10.3390/s20041011
25. Ramezanifar H, Ghazvini M, Shojaei M. A new data aggregation approach for WSNs based on open pits mining. *Wirel. netw*. 2021; 27(1). DOI: 10.1007/s11276-020-02442-9.
26. Shobana M, Sabitha R, Karthik S. Cluster-based systematic data aggregation model (CSDAM) for real-time data processing in large-scale WSN. *Wirel Pers Commun*. 2021;117(4):1-9. DOI: 10.1007/s11277-020-07054-2.
27. Idrees AK, Al-Qurabat AKM. Energy-Efficient Data Transmission and Aggregation Protocol in Periodic Sensor Networks Based Fog Computing. *J. Netw. Syst. Manag*. 2021; 29(1):1-24. DOI: 10.1007/s10922-020-09567-4
28. Al-Qurabat AKM, Abou Jaoude C, Idrees AK. Two tier data reduction technique for reducing data transmission in IoT sensors. In: *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) 2019*. 2019 (pp. 168-173). DOI: 10.1109/IWCMC.2019.8766590
29. Idrees AK, Abou Jaoude C, Al-Qurabat AKM. Data Reduction and Cleaning Approach for Energy-saving in Wireless Sensors Networks of IoT. In: *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2020*. 2020 (pp. 1-6). DOI: 10.1109/WiMob50308.2020.9253429
30. Al-Qurabat AKM, Idrees AK, Abou Jaoude C. Dictionary-Based DPCM Method for Compressing IoT Big Data. In: *2020 International Wireless Communications and Mobile Computing, IWCMC.* 2020 (pp. 1290-1295). DOI: 10.1109/IWCMC48107.2020.9148492
31. Jawad GA, Al-Qurabat AKM, Idrees AK. Compression-based Block Truncation Coding technique to Enhance the Lifetime of the Underwater Wireless Sensor Networks. *IOP Conf Ser Mater Sci Eng*. 2020; 928(3):032005. DOI: 10.1088/1757-899X/928/3/032005
32. Al-Qurabat AKM, Idrees AK. Distributed data aggregation protocol for improving lifetime of wireless sensor networks. *Qalaai Zanist Sci. J*. 2017; 2(2):204-15. DOI: 10.25212/lfu.qzj.2.2.22
33. Abdulzahra SA, Al-Qurabat AKM, Idrees AK. Compression-based Data Reduction Technique for IoT Sensor Networks. *Baghdad Sci. J*. 2021; 18(1):0184. DOI: 10.21123/bsj.2021.18.1.0184
34. Fu TC, Chung FL, Ng CM. Financial Time Series Segmentation based on Specialized Binary Tree Representation. DMIN. 2006; 2006:26-9.
35. Zhang Z, Jiang J, Wang H. A new segmentation algorithm to stock time series based on pip approach. In: *2007 International Conference on Wireless Communications, Networking and Mobile Computing 2007*. 2007 (pp. 5609-5612). DOI: 10.1109/WICOM.2007.1374
36. Jiménez P, Nogal M, Caulfield B, Pilla F. Perceptually important points of mobility patterns to characterise bike sharing systems: The Dublin case. *J. Transp. Geogr*. 2016; 54:228-39. DOI: 10.1016/j.jtrangeo.2016.06.010
37. Bodik P. Intel berkeley research lab. 2004, Accessed on: Jan. 1, 2021, [Online] Available: http://db.csail.mit.edu/labdata/labdata.html.

طريقة تجميع البيانات المستندة إلى النقاط المهمة إدراكيًا لشبكات أجهزة الاستشعار اللاسلكية

ايمان داخل عيدان سعيدي **علي كاظم محمد الغرابي***

قسم علوم الحاسوب، كلية العلوم للبنات، جامعة بابل، بابل، العراق.

الخلاصة:
يستهلك إرسال واستقبال البيانات معظم الموارد في شبكات الاستشعار اللاسلكية (WSNs). تعد الطاقة التي توفرها البطارية أهم مورد يؤثر على عمر WSN في عقدة المستشعر. لذلك، نظرًا لأن عُقد المستشعر تعمل بالاعتماد على بطاريتها المحدودة ، فإن توفير الطاقة ضروري. يمكن تعريف تجميع البيانات كإجراء مطبق للقضاء على عمليات الإرسال الزائدة عن الحاجة ، ويوفر معلومات مدمجة إلى المحطات الأساسية ، مما يؤدي بدوره إلى تحسين فعالية الطاقة وزيادة عمر الشبكات اللاسلكية ذات للطاقة المحدودة. في هذا البحث ، تم اقتراح طريقة تجميع البيانات المستندة إلى النقاط المهمة إدراكيًا (PIP-DA) لشبكات المستشعرات اللاسلكية لتقليل البيانات الزائدة عن الحاجة قبل إرسالها إلى المحطة الاساسية. من خلال استخدام مجموعة بيانات Intel Berkeley Research Lab (IBRL) ، تم قياس كفاءة الطريقة المقترحة. توضح النتائج التجريبية فوائد الطريقة المقترحة حيث تعمل على تقليل الحمل على مستوى عقدة الاستشعار حتى 1.25٪ في البيانات المتبقية وتقليل استهلاك الطاقة حتى 93٪ مقارنة ببروتوكولات PFF و ATP.

الكلمات المفتاحية: تجميع البيانات، توفير الطاقة، النقاط المهمة أدراكياً، شبكة المستشعرات اللاسلكية.