EDaTAD: Energy-aware Data Transmission Approach with Decision-Making for Fog Computing-based IoT Applications

Ali Kadhum Idrees *[1], Tara Ali-Yahiya[2], Sara Kadhum Idrees[1,3], and Raphael Couturier[4]

[1] Dept. of Information Networks, College of Information Technology, University of Babylon, Hilla, 51001, Babylon, Iraq.
[2]Laboratoire Interdisciplinaire des Sciences du Numérique, CNRS, Université Paris-Saclay, Street, Gif-sur-Yvette, 91190,France.
[3]CIAD (UMR 7533), Université de Technologie de Belfort Montbéliard, UTBM, Belfort, France.
[4]FEMTO-ST Institute/CNRS, Université de Franche-Comté, Belfort, France.

2024

Abstract

In the fog computing-based Internet of Things (IoT) architecture, the sensor devices represent the basic elements needed to sense the surrounding environment. They gather and send a huge amount of data to the fog gateway and then to the cloud due to their use in various real-world IoT applications. This would lead to high data traffic, increased energy consumption, and slow decisions at the fog gateway. Therefore, it is important to reduce the transmitted data to save energy and provide an accurate decision regarding the safety and health of the building's environment. This paper suggests an energy-aware data transmission approach with decision-making (EDaTAD) for Fog Computing-based IoT applications. It works on two-level nodes in the fog computing-based TI architecture: sensor devices and fog gateways. The EDaTAD implements a Lightweight Redundant Data Removing (LiReDaR) algorithm at the sensor device level to lower the gathered data before sending it to the fog gateway. In the fog gateway, a decision-making model is proposed to provide suitable decisions to the monitoring staff in remote monitoring applications. Finally, it executes a Data Set Redundancy Elimination (DaSeRE) approach to discard the repetitive data sets before sending them to the cloud for archiving and further analysis. EDaTAD outperforms other methods in terms of transmitted data, energy consumption, and data accuracy. Furthermore, it assesses the risk efficiently and provides suitable decisions while decreasing the latency time.

Keywords: Internet of Things, Transmission Data Reduction , Decision-making, Clustering, Fog Computing, Energy-efficiency.

1 Introduction

Nowadays, connected sensor devices outnumber the population. In daily life, several sensors collect data for various applications. Data collection, surveillance, and sensing have recently been introduced in several applications such as road traffic, transport, environment, smart grid, military, healthcare, smart buildings, remote education, smart cities, environment, etc. [18]. Depending on the application needs, these sensors can capture data in the form of numbers, pictures, audio, or video. The smart wireless devices which generate data represent the essential component in the fog computing-based IoT architecture as depicted in Figure 1.

*Corresponding author: ali.idrees@uobabylon.edu.iq

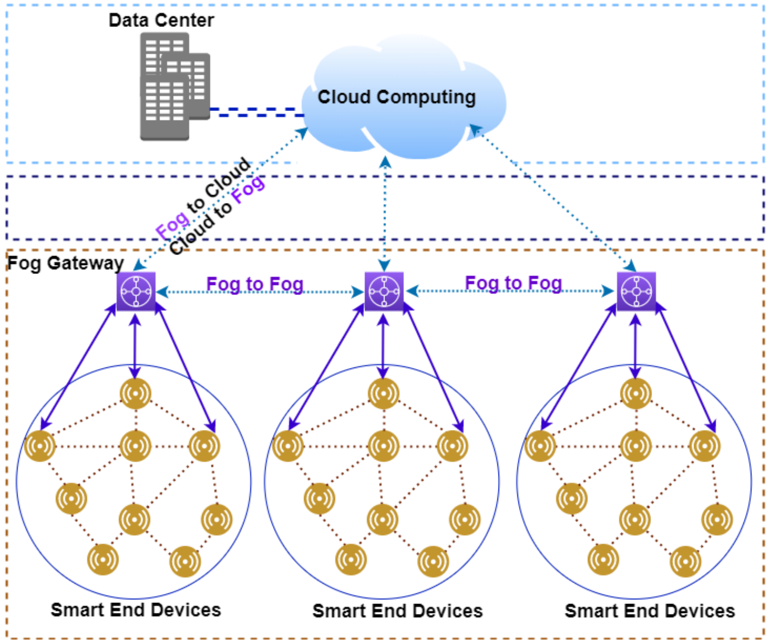


Figure 1: The architecture of the fog computing based IoT.

In the fog computing-based IoT network, the redundancy of data transmission generated from the spatial and temporal similarity represents a major challenge [15]. The increased volume of transferred data leads to increased overhead and response delay on the network, as well as the increased power consumption of smart devices and fog gateways. To lower data transmission latency and mitigate internet congestion, data transmission reduction has been approved as a significant element of the fog computing-based IoT network [12]. To minimize the amount of data transferred, conserve energy on these devices, and decrease the data transmission latency while keeping a sufficient ratio of data accuracy. it is crucial to eliminate the repetitive data at both the sensor and fog devices. The accuracy of the data refers to the ratio of lost data. It is calculated by counting the number of lost readings in the received data at the gateway after applying our approach at the sensor nodes. Additionally, the reduction of data using EDaTAD based LiReDaR algorithm at the sensor nodes can minimize the transmitted data to the Fog gateway, which introduces low latency and enables quick decisions [1]. This is a crucial measurement for the fog computing-based IoT applications. A decision-making model is based on environmental parameters such as temperature, humidity, and light in the building to evaluate the global risk and determine the level of safety and health of the building's environment. Temperature control is important for preserving a comfortable and safe indoor environment, as high temperatures can pose health risks. Maintaining optimal indoor humidity levels (between 30%-60%) is crucial for indoor air quality and occupant health, as high humidity can trigger allergies, respiratory problems, and asthma symptoms. Moreover, improper lighting influences circadian rhythms, mood, and productivity [4, 21, 25, 28, 31]. The following contributions are introduced in this paper:

- An Energy-aware Data Transmission Approach with Decision-making (EDaTAD) for Fog Computing-based IoT applications is proposed. The EDaTAD works on two-level devices of the fog computing-based IoT architecture: sensor devices and fog gateway. The purpose of the EDaTAD scheme is to minimize the volume of sent data for both smart devices and gateway to save energy and reduce the delay of data transmission whilst also maintaining a sufficient level of data accuracy. Moreover, the reduction of data by the EDaTAD can introduce low latency and enable quick decisions regarding the global risk of safety and health of the environment of the building.

- A Lightweight Redundant Data Removing (LiReDaR) Algorithm is suggested and implemented at the level of sensor nodes by EDaTAD to eliminate the similar data and then transmitting them to the fog gateway.

- A novel decision-making model is suggested to evaluate the global risk of safety and health to the environment of the building and to provide an appropriate decision to the monitoring staff. This decision is an indication of how the environment is safe and healthy for the people in the building. This model is executed by EDaTAD at the fog gateway (i.e., close sensor nodes) to provide a fast response in the case of an emergency in remote monitoring applications.

- A Data Set Redundancy Elimination (DaSeRE) Algorithm is proposed and implemented by EDaTAD at the fog gateway to discard the repetitive data reading sets that collected from sensor devices before forwarding them to the data centers of the cloud for archiving and more analysis.

- Extensive simulation studies by OMNeT++ simulator are achieved to ensure the efficiency of the proposed EDaTAD approach. Real data are used during the simulation from smart sensor nodes deployed at the Intel Berkeley Lab. The EDaTAD approach is compared to some techniques such as Harb method [8], PFF scheme [3], and ATP technique [7]. It can be observed from the results that the suggested EDaTAD approach improves better performance compared to other techniques, evaluates the risk of a monitored area of interest efficiently and provides an accurate decision to the monitoring staff while decreasing the latency time.

This paper can be orgnized as follows: The related work is introduced in the next section. In section III, the suggested EDaTAD approach is introduced. Section IV presents the performance evaluation of the proposed EDaTAD approach. Section V explores the conclusion and perspectives.

2 Related Work

One of the major challenges in the fog computing based IoT system is to lower the massive amount of sent data, preserve some energy, and decrease the data sending latency while keeping the accuracy of the obtained data at the final level of the network at a suitable level. Researchers have come up with several data-reduction techniques in the modern era to lower network transmission costs. This section addresses some related works approaches that focus on this topic. There are three main approaches focused on reducing the collected data in the network such as data compression [2, 15, 22] and data reduction [3, 7–9, 12, 13], and data sampling and reduction [6, 11, 14, 16, 17, 24]. In the compression based approaches, the authors in [5] join the lossy compression and the discrete cosine transform and then perform the quantization for the Huffman coding to achieve a large ratio of compression with sufficient distortion in the original signal. In [2], the authors use the Discrete Wavelet Transform with lifting scheme to compress the data in the WBSNs (Wireless Body Sensor Networks). Then, they combine lossy Lightweight Temporal Compression algorithms with the lossless Differential Pulse Code Modulation to improve the compression ratio and decrease the error rate of data construction. Rajasekar and Pushpalatha [22] suggested a lossless compression technique using Huffman based discrete cosine transform to decrease the complexity of data and to enhance data privacy. In [15], The authors combined Huffman encoding and clustering, two effective techniques that the authors used to propose a lossless compression method. The information is organised into groups. Then, Huffman encoding is used to compress each group of data. In the data reduction based approaches, the Prefix-Frequency Filtering (PFF) method in [3] operated on both sensor and aggregator nodes. They utilize Jaccard similarity in the sensor node to eliminate duplicate data and set similarity in the gateway to decrease duplicated data sets. The authors present a novel prefix filtering technique for analysing the similarity of sets in sensor networks. A frequency filtering technique is suggested, which takes advantage of the structure of measurements based on their frequencies. The frequency of a given measure is determined by the quantity with which it occurs within a given set. This approach consists of two distinct phases. In the initial phase, each node performs the compacting of its measurement set using a link function. The second is specified at the level of the aggregator, which is where the frequency filtering method is implemented. This strategy is improved by the newly developed techniques in [9] that reduce data redundancy in the sensed data before sending it to the gateway. In [7],the researchers explain an ATP (Aggregation and Transmission Protocol) approach to minimize the amount of captured data at the sensor device before transferring it to the sink. In this study, they provide a novel adaptive protocol that operates independently on individual sensor nodes. The primary objective of this protocol is to minimise data transfer and conserve energy resources. A cluster-based strategy is achieved in this study, wherein data is periodically sent from sensor nodes to their respective cluster heads (CHs). During the aggregation step, the suggested method aims to identify commonalities among data collected during a specific period p, with the objective of removing data redundancy from the raw data. The sensor node conducts periodic correlation analysis of data during the transmission phase, employing a one-way analysis of variance (ANOVA) model and the Fisher test. Harb et al. [8] proposed a two-level approach of data reduction to save power and maximize the lifetime of the network. The sensor device cleans the collected data to lower the size of data readings before transferring it to the cluster head. The cluster head implements the sets-similarity algorithm to eliminate duplicate data sets before delivering them to the next level of the network. The authors in [3, 7–9] define the data accuracy as the ratio of lost data. It is calculated by counting the number of lost readings in the received data at the gateway after applying our approach at the sensor nodes.

The work in [27] proposed a data reduction method based on the spatial-temporal correlation of data to save energy in sensor networks. The presented approach reduces the volume of transmitted data while maintaining data accuracy. The works in [30] focused on achieving the adaptive data sampling to remove the data redundancy in the gathered data. The correlation between the two consecutive data readings vectors is exploited to lower the size of data readings before forwarding them to the next level of the network. The authors in [12] proposed power saving data transmission and reduction for IoT network based fog computing. They proposed a protocol works on two levels: sensor node and fog node. At the sensor device, they presented a method that combines between the grouping and encoding to discard the repetitive data and then sends them to the fog gateway. Then, they applied a grouping method based on the DTW (Dynamic Time Warping) to discard the repetitive data sets collected at the fog gateway from the IoT devices and then transmit them to the cloud data center.

For the data sampling and reduction-based methods, in [14], the authors proposed a data sampling algorithm that works in an adaptive and distributed way to reduce the data and save the energy in a periodic sensor network. Their proposed method applied three phases in each round. The first phase uses the Map reduce to remove redundant data before sending them to the next level of the network. The sampling rate adaptation is done at the end of each round using the longest common subsequence similarity and grouping method. In [17], the authors proposed adaptive data sampling method for periodic sensor networks. They implemented the data aggregation at each sensor node to decrease the redundant data before sending them to the cluster head. Then, the data sampling algorithm is executed at the cluster head to produce the new sampling rate based on the collected data from the sensor nodes. The authors in [6] proposed a two-fold method: first, they proposed a data sending model based on the spearman coefficient and clustering to reduce transmitted data to the sink and save the energy of sensor nodes. Second, they proposed a method to adapt the sampling rate of the sensor nodes based on the similarity between their collected data at the sink. In [11], the authors proposed a three phases approach called All-in-one. They applied either compression or aggregation at the sensor nodes to reduce the transmitted data to the cluster head in the first phase. They propose a sampling algorithm to adapt the sampling rate based on the variation of the monitored region. Finally, they applied at the cluster head a data reduction based on the spatial correlation of collected data of the sensor nodes using the clustering methods to reduce the collected data before sending them to the sink. In [16, 24], the authors proposed an energy-efficient algorithm for sampling rate adaptation of biosensor nodes in the wireless body sensor network. They first removed the data redundancy before sending them to the aggregator and then adapting the sampling rate according to the situation of the patient. Table 1 refers to the features of existing works.

Despite illustrating different methods for data reduction, to the best of our knowledge, an efficient approach that integrates lowering the size of data on the sensor nodes and giving a fast decision on the fog node about the situation of safety and health of the building environment in the applications for remote surveillance is not found. For example, the presented related data reduction techniques can decrease the accuracy of the data, and therefore they do not ensure an accurate risk evaluation and decision because they did not use a decision model based on this low-quality data. In addition, these methods were not employed at the fog gateway to achieve the decision-making to monitor the safety and health of the building environment remotely based on the received data from the sensor nodes. Moreover, these related works did not consider the latency time in their studies.

An Energy-aware Data Transmission Approach with Decision-making (EDaTAD) in fog computing based TI applications is proposed. To close this research gap, our contribution is threefold: (1) A Lightweight

Table 1: The features of existing works.

Ref No.	Compression	Redundancy deletion	Aggregation	Level of operation	Data sampling	Decision-making
[2]	Yes	Yes	No	Sensor	No	No
[22]	Yes	Yes	No	Sensor	No	No
[15]	Yes	Yes	No	Fog	No	No
[3]	No	Yes	Yes	Sensor, Gateway	No	No
[9]	No	Yes	Yes	Sensor, Gateway	No	No
[7]	No	Yes	Yes	Sensor, Gateway	No	No
[8]	No	Yes	Yes	Sensor, Gateway	No	No
[27]	No	Yes	Yes	Sensor, Gateway	No	No
[30]	No	Yes	No	Sensor	Yes	No
[12]	Yes	Yes	No	Sensor, Fog	No	No
[14]	No	Yes	No	Sensor	Yes	No
[17]	No	Yes	Yes	Sensor	Yes	No
[6]	No	Yes	No	Sensor	Yes	No
[11]	Yes	Yes	Yes	Sensor, Gateway	Yes	No
[24]	No	Yes	No	Sensor, Gateway	Yes	No
[16]	No	Yes	Yes	Sensor, Gateway	Yes	Yes

Redundant Data Removing (LiReDaR) algorithm is suggested and implemented on the level of sensor nodes by the EDaTAD approach to eliminate the redundant data and then transmit them to the fog gateway. (2) Using the received data from the sensor nodes, we propose a decision-making model to remotely monitor the safety and health of the building environment and to provide a suitable and fast decisions to the monitoring staff with a reduced latency time. This model is implemented by EDaTAD at the fog gateway (i.e., close sensor nodes) to provide a quick response in the case of an emergency in remote monitoring applications. (3) A Data Set Redundancy Elimination (DaSeRE) Algorithm is suggested and executed by EDaTAD at the fog gateway to further eliminate the spatial data redundancy of the received data sets of sensor devices before forwarding them to the cloud for additional processing and archiving.

3 The EDaTAD approach

The EDaTAD approach is illustrated in more detail in this section. The EDaTAD approach is shown in Figure 2 . Three algorithms are used by the EDaTAD approach at both sensor and fog nodes. It operates into periods. The network's overall periods determine how long it will stay alive. These algorithms' main objectives are to eliminate duplicate data before sending them to the next level of the network, to remotely monitor the environment's health and safety, and to give the monitoring staff a suitable decision. It lowers the cost of transmitting/receiving, increases lifetime, and speeds up the response to monitoring staff whilst also maintaining a sufficient level of data accuracy.

3.1 Data processing at sensor device

The Lab building environment in the area of interest is sensed by sensor devices periodically. The period includes several slots, and the sensor node catches a data measurement in each slot of time. The gathered data measurements throughout the period comprise the set of data measurements $\delta = \{s_1, \ldots, s_\rho\}$, where ρ refers to the size of data in the period. The δ is the set of collected data from a single sensor (not multiple sensors. The data readings that were sensed during one period and represented by the data set δ tend to be identical or very close, especially in cases where the monitored environment has not changed for a long time. Hence, it is necessary to remove this increased number of redundant measurements from the set of data measurements δ at this smart device level. Therefore, a Lightweight Redundant Data Removing (LiReDaR) Algorithm is suggested and applied at the sensor node by the EDaTAD approach to eliminate the repetitive

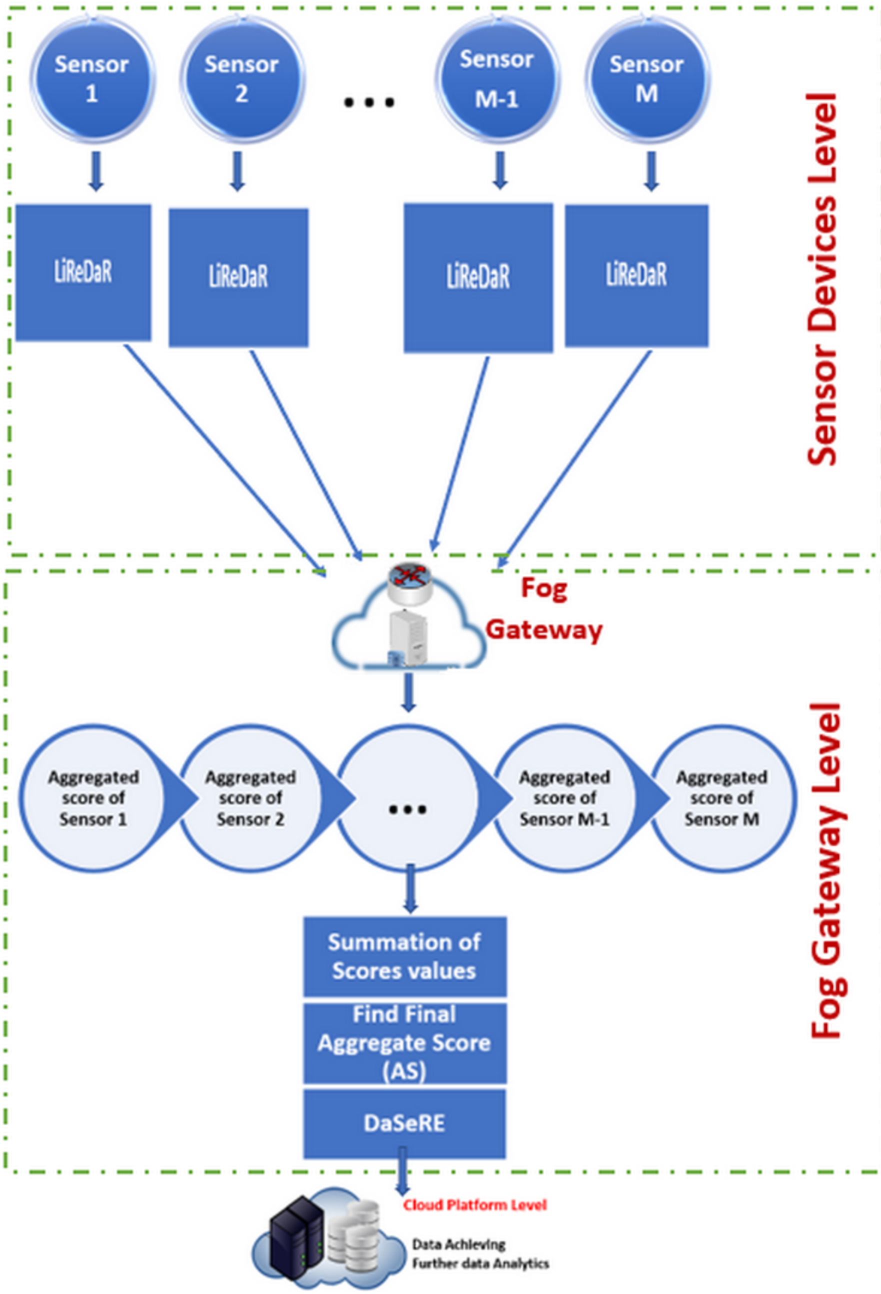


Figure 2: The EDaTAD approach.

data and then forward them to the fog gateway. Algorithm 1 illustrates the LiReDaR algorithm.

In algorithm 1, steps 1-17 build a new set of reduced data measurements after eliminating the data measurement redundancy from the set of sensed data measurements δ. The set of reduced data measurements β starts with an empty list, then the first measurement of δ is assigned to β. After that, the comparison is done for each measurement δ_k ($2 \leqslant$ k $\leqslant \rho$) with the data measurements in β according to a particular threshold α specified by the application. If the condition of comparison is satisfied (difference between δ_k and $\beta \leq \alpha$), it means that the δ_k is similar to one of the data measurements in β and it will be deleted. Otherwise, the data measurement will be appended to the set β. The function Size() returns the size of the list. The threshold α refers to the highest acceptable difference between two measurements to be similar. If the difference between two measurements is less than or equal to α, then they are considered similar. This

Algorithm 1 *LiReDaR Algorithm*

Require: δ:*Set of data measurements*, ρ: *Number of data measurements*, α: *threshold*
Ensure: E: *the compressed data of reduced vector*

1: $\beta \leftarrow \phi$;
2: $\beta \leftarrow \beta \cup \delta_1$;
3: **for** $k \leftarrow 2$ to ρ **do**
4:     $i \leftarrow 1$;
5:     **while** $i \leq Size(\beta)$ **do**
6:         $\varphi \leftarrow |\delta_k - \beta_i|$;
7:         **if** $(\varphi > \alpha)$ **then**
8:             $i \leftarrow i + 1$;
9:         **else**
10:             $i \leftarrow Size(\beta) + 1$;
11:         **end if**
12:     **end while**
13:     **if** $(\varphi > \alpha)$ **then**
14:         $\beta \leftarrow \beta \cup \delta_k$;
15:     **end if**
16: **end for**
17: $E \leftarrow \phi$;
18: $E \leftarrow E \cup COMPRESSOR(Size(\beta))$;
19: **for** $j \leftarrow 1$ to $Size(\beta)$ **do**
20:     $E \leftarrow E \cup COMPRESSOR(\beta_j)$;
21: **end for**
22: return E;

threshold is determined according to the application used. If the application is critical to the data loss, it is set to the minimum value; otherwise, it can be set to an acceptable value that balances between the data reduction ratio and accuracy. In this paper, to make a fair comparison, the threshold α is set to the same values as in the papers with which we compared our work. The difference function is defined as the difference between two data measurements s_i and $s_j \in \delta$, which are collected by the sensor device. The two readings are not considered similar if the difference between s_i and s_j is greater than α, otherwise, they are similar. In algorithm 1, the parameter φ is the difference between two measurements. The application's user will determine the α threshold.

Steps 18-22 are in charge of encoding the resulting set β of reduced data at the sensor device using a simple encoding method. Every data measurement in β is compressed in two bytes using the COMPRESSOR function. Then, the compressed data measurement is added to the file E. Figure 3 refers to the 16-bits representation of data measurement. The sign bit utilizes the binary values "1" and "0" for the negative and positive numbers respectively. The integer part is the next part of the data measurement representation that takes 8-bits. The last part takes 7-bits that represents the fraction part of the data measurement.

Figure 3: The 16-bits representation of data measurement.

The COMPRESSOR function plays an important role in reducing the size of the reduced data set β by using a simple encoding method. This results in further lowering the volume of data, saving energy, and maintaining a suitable quality of data. The compressed file E is forwarded to the fog gateway. The time requirement of Algorithm 1 is O(ρ $Size(\beta)$) . The storage requirement is O(ρ + $Size(\beta)$). The Compressor function in Algorithm 1 needs time requirement equal to O(log n), where n represents the given input digits number. Figure 4 shows an illustrative example about applying Algorithm 1.

Suppose we have the following input parameters in Algorithm 1:

δ = {19.98, 19.98, 19.98, 19.30, 19.30}, ρ= 5, α= 0.05

β= [19.98]

K= 2 After applying Steps 4-12 φ = 0 0 > 0.05? No	K= 3 After applying Steps 4-12 φ = 0 0 > 0.05? No
K= 4 φ = 0.68 0.68 > 0.05? Yes β= [19.98, 19.30]	K=5 After applying Steps 4-12 φ = 0 0 > 0.05? No

Hence, the reduced set that must be transmitted is β= [19.98, 19.30], Size(β) =2

After applying steps 17-21, the transmitted file E will be as follow:

1	0000001	00000000	1	00010011	1100010	1	00010011	0011110

Figure 4: Illustrative example about applying Algorithm 1.

3.2 Data processing at fog gateway:

Using the received data from the sensor nodes, the EDaTAD approach will use this data by two proposed algorithms: (1) A decision-making model to remotely monitor the safety and health of the building environment and to provide a suitable and fast decision with reduced latency time to the monitoring staff. (2) A Data Set Redundancy Elimination (DaSeRE) Algorithm to further remove the data redundancy of the received data sets before sending them to the cloud for additional analysis and archiving.

3.2.1 Decision Making Model

The main goal of deploying the sensor devices in the area of interest is to collect the data and then achieve decision making based on the collected data. This section introduces the proposed decision-making model that enables the monitoring staff to take the appropriate decision according to the situation of the monitored area of interest. This model implemented by the EDaTAD approach. The proposed decision-making model can be adapted to other applications of sensor networks based on the application requirements. The principal idea of the proposed model is inspired from the National Early Warning Score (NEWS) with some modifications, where the NEWS uses the physiological scoring system to assess the situation of the patients and provide the suitable treatment and care for them in the case of a risk [26]. Table shows the proposed Lab Environment Early Warning Score (LEEWS).

Table 2: Lab Environment Early Warning Score.

Lab. Environment	Scores						
	3	2	1	0	1	2	3
Temperature	<= T1	T2 -T3	T4-T5	T6-T7	T8-T9	T10-T11	>T12
Humidity	H1-H2	H3-H4	H5-H6	H7-H8	H9-H10	H11-H12	H13-H14
Light	<L1	L2-L3	L4-L5	L6-L7	L8-L9	L10-L11	L12–L13

In Table 2, the values of the parameters T1,..., T12; H1,..., H14; and L1,..., L13 are assigned by the decision maker according to the application requirements. Table 3 introduces the LEEWS parameters' values setting that are used in this paper.

The proposed decision-making model is based on three tables: the Lab Environment Early Warning Score (LEEWS), a chart monitoring staff response, and the aggregated score, safety risk level, and the decision type. In Table 3, the range of values for the lab environment parameters such as temperature, humidity, and light that are assigned to score "0" refers to the normal state of the environment that will be comfortable and healthy [4, 21, 25, 31].

Table 3: LEEWS parameters' values setting.

Lab. Environment	Scores						
	3	2	1	0	1	2	3
Temperature	<= 0	1 -10	11-19	20-22	23-30	31-40	>40
Humidity	0-10	11-20	21-29	30-60	61-70	71-80	81-100
Light	<50	50-150	151-299	300-500	501-1000	1001-10000	10001–25000

The range of values of the same parameters to the other scores 1, 2, and 3 are custom assigned and it can be re-changed by the decision-maker according to the application used. Table 4 introduces the chart monitoring staff response. The reading scores of the sensor nodes in the building are aggregated based on LEEWS. The obtained value of the aggregation score will be matched with the range of score values in Table 5 to determine the risk level and the corresponding type of decision, while it can take the correct action regarding the score values according to the chart monitoring staff response (see Table 4). This chart is categorized into four classes according to the severity level of environment condition, beginning with the normal situation and graduating to high risk of safety and health situations. The range of parameter values in Tables 4 and 5 can be modified by the expertise of the monitoring staff according to the application requirements.

Table 4: The chart monitoring staff response..

LEEWS Score	Frequency of Monitoring	Monitoring Staff Response
0	Minimum 24 hourly	Comfort: continue the monitoring normally
Total: 1-2	Minimum4-8 hourly	Inform the technical staff to visit the building and to see if some action required
Total: 3 Or 4-6	Increase the monitoring frequency to a minimum of 1 hourly	Risk for safety and health: Urgently inform the technical staff to visit the building and take the appropriate action for reaching the comfortable state in the building.
Total: 7 or more	continuous monitoring of environment sensor readings	High risk for safety and health: Immediately inform the technical staff to visit the building and take the appropriate fast action for reaching the comfortable and healthy state in the building.

The value of the global risk of safety and health represents an indication of how the environment of the building is safe and healthy for the people in the Lab. This value is computed in the fog gateway by aggregating the scores of three sensor nodes with different environment parameters (temperature, humidity, and light). The Fog gateway receives the readings of a period and then aggregates the scores to make the appropriate decision. The aggregated score can be calculated as follows.

$$T^j \leftarrow \frac{\sum_{i=1}^{\delta^j} S_i^t}{\delta^j}. \tag{1}$$

$$H^j \leftarrow \frac{\sum_{i=1}^{\beta^j} S_i^h}{\beta^j}. \tag{2}$$

$$L^j \leftarrow \frac{\sum_{i=1}^{\gamma^j} S_i^l}{\gamma^j}. \tag{3}$$

The T^j, H^j, and L^j refer to the average value of the readings scores of sensor node j that was received by the fog gateway for the temperature, humidity, and light readings respectively. The S_i^t, S_i^h, and S_i^l are the i^{th} score value for the temperature, humidity, and light readings respectively. The δ^j, β^j, and γ^j are the number of score values of sensor node j for the temperature, humidity, and light readings respectively. In this paper, the fog gateway aggregates the summation of score values (T^j, H^j, and L^j) for each sensor node at the end of each period to calculate the final Aggregated Score (AS) value that can be computed as follows:

$$Sum_j \leftarrow T_j + H_j + L_j, j = 1, ..., M \tag{4}$$

$$AS \leftarrow \max_{j \in M} Sum_j \tag{5}$$

Where M refers to the number of sensor devices. In order to calculate the safety risk level and the required decision for the monitored building, the average aggregated score values calculated at the end of each period is matched with the aggregated score values in Table 5. This can determine the accurate safety risk level and the corresponding decision type that should be taken by the monitoring staff. Table 5 refers to the aggregated score, safety risk level, and the decision type.

Table 5: The aggregated score, safety risk level, and the decision type.

Aggregated Score	Safety Risk Level	Decision Type
<1	Low	D1
1-2.99	Low-Medium	D2
3-3.99	Medium	D3
4-6.99	Medium-High	D4
>= 7	High	D5

3.2.2 DaSeRE Algorithm

The focus of this section is on the main role of the EDaTAD approach in reducing the received redundant data measurements vectors from the nodes in the first level in the network. It will execute the proposed Data Set Redundancy Elimination (DaSeRE) Algorithm to eliminate the redundant data set resulted from the spatial correlation between the received data sets of sensor devices. The DaSeRE algorithm of EDaTAD approach is responsible for removing the unnecessary data sets at the fog gateway before delivering them to the sink. The DaSeRE algorithm first decodes the received encoded data of each sensor node j using the DECOMPRESSOR function, and saves it in vector $X^j = \{x_1^j, x_2^j, ..., x_T^j\}$, where T is the total number of data measurements of sensor device j. Then, the variance of each data measurement set X^j is computed. The result is a set of variance values expressed as $R \leftarrow \{r_1, r_2, ..., r_M\}$.

The DaSeRE algorithm uses the Mini-batch k-Means (see Algorithm 2) to cluster the data measurements sets of the sensor devices according to their variance values into groups of data measurement sets. This algorithm is the modified version of the k-means algorithm and it reduces the processing time in the large datasets by using mini-batches. Hence, it is faster than the k-means algorithm and produces better solutions [20, 23]. Algorithm 2 requires $O(T.(b + K))$ of time complexity.

One data measurement set is elected as a representative set for each group. The results are a group of data measurement sets, each one represents at least one or more of the data measurements sets of sensor devices. Algorithm 3 refers to the DaSeRE Algorithm.

Algorithm 2 *Mini-batch k-Means clustering*

Require: K:*number of clusters, b: size of mini-batch,* Y: *dataset, M: size of dataset*
Ensure: C: *set of centroids*

for $j \leftarrow 1$ to K **do**
  $C_j \leftarrow Y_i$; $//randomlyselectY_ifromY(1 \leq i \geq M)$
  $S_j \leftarrow Y_i$;
  $N_j \leftarrow 1$;
end for
while $Stoping\ Creterion\ is\ not\ satisfied$ **do**
  $L \leftarrow b\ samples\ selected\ randomly\ from\ Y$;
  for $Each\ i \in L$ **do**
    $A_i \leftarrow \text{argmin}_{j \in K} ||Y_i - C_j||$;
    $S_{A_i} \leftarrow S_{A_i} + Y_i$;
    $N_{A_i} \leftarrow N_{A_i} + 1$;
  end for
  for $Each\ i \in K$ **do**
    $C_j \leftarrow S_j / N_j$;
  end for
end while
return C;

Algorithm 3 *DaSeRE Algorithm*

Require: K:*number of clusters, b: size of mini-batch, D: encoded reduced vectors, M:number of reduced vectors in D*
Ensure: EDS: *Encoded data reading sets*

1: **for** $j \leftarrow 1$ to M **do**
2:   $X^j \leftarrow DECOMPRESSOR(D^j)$;
3:   $R^j \leftarrow Variance(X^j)$;
4: **end for**
5: $C \leftarrow Mini-batch\ k-Means\ clustering(K, b, R, M)$;
6: $EDS \leftarrow \phi$;
7: $EDS \leftarrow EDS \cup COMPRESSOR(K)$;
8: **for** $Each\ i \in K$ **do**
9:   $DS^i \leftarrow BringRepresentativeDataSet()$;
10:   $EDS \leftarrow EDS \cup COMPRESSOR(Length(DS^i))$;
11:   **for** $k \leftarrow 1$ to $Size(DS^i)$ **do**
12:     $EDS \leftarrow EDS \cup COMPRESSOR(DS^i_k)$;
13:   **end for**
14: **end for**
15: return EDS;

In this algorithm, the function DECOMPRESSOR() is responsible for decoding the received file D^j of sensor device j into a set of data readings X. The function COMPRESSOR() is responsible for encoding each reading (see Figure 3) into 16 bits for appending each one to the encoded file before transmitting it to the next level of the network. The function BringRepresentativeDataSet() brings the representative data set according to its centroid. The fog gateway sends the encoded file of the representative data sets with the identifications (ids) of sensor devices to the next level of the network. The DaSeRE Algorithm removes the repetitive data reading sets to reduce the volume of the transmitted data, save energy, reduce the latency, and preserves the accuracy with an acceptable level. Moreover, it further reduces the transmitted data by encoding them before sending them to the next level of the network. In DaSeRE Algorithm, steps 1-4 require time complexity for $O(M.Len(X))$. The time complexity of step 5 is $O(T.(b+K))$ while the steps 6-14 require $O(K.Len(DS))$. Hence, the time requirement for the algorithm 3 is $O(MAX((T.(b+K)), (M.Len(X)), (K.Len(DS)))$, where

$Len(X)$ and $Len(DS)$ refer to the lengths of vectors X and DS respectively.

4 Performance Evaluation

This section introduces the performance assessment of the proposed EDaTAD approach using a network simulator called OMNeT++ [29]. Many simulation experimental results have been conducted using real sensed data measurements from the sensor nodes which were distributed in the Lab of the Intel Berkeley [19]. This Lab includes 47 devices that sense the readings (e.g., light, temperature, voltage, and humidity) every 31 seconds from the laboratory building environment. In our experiments, we consider a period of 20, 50, 100, 200, 500, and 1000 readings. Since the sensor nodes in the Intel Berkeley Lab sense one reading every 31 seconds, we can suppose that the long of the period is ρ * 31 seconds, where ρ refers to the size of data in the period. The experiments regarding the sensor level consider ρ equal to 20, 50, and 100 readings, while experiments regarding the fog gateway consider ρ equal to 200, 500, and 1000 readings. Hence, the sensors continue sense ρ readings each period until the end of the dataset of Intel Berkeley Lab. The includes a log of about 2.3 million readings collected from these sensors. The dataset is about 2.3 million readings $nearly150MB$ sensed from these sensor devices. In these simulation experiments and for the sake of simplicity, the EDaTAD approach only uses the temperature data measurements. Figure 5 illustrates the network architecture used during this simulation.

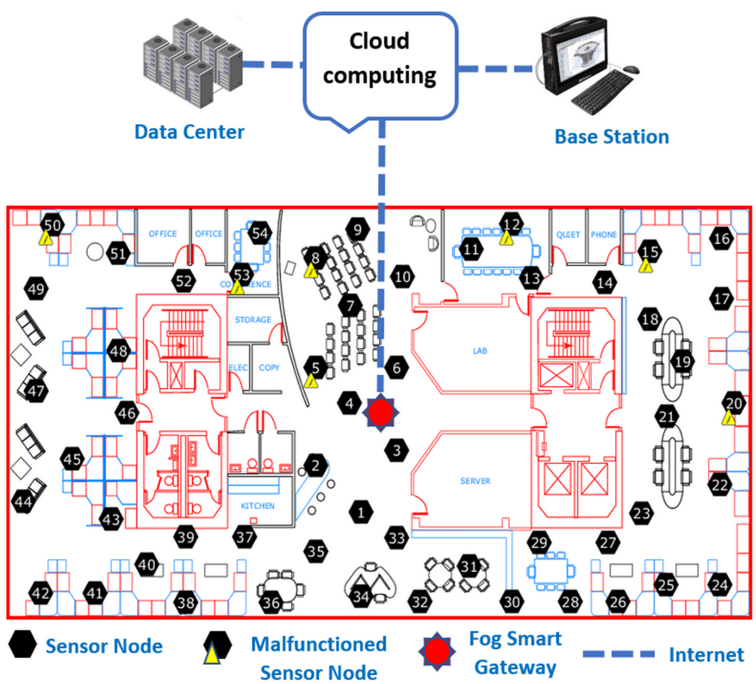


Figure 5: The sensor network of Intel Berkeley Lab.

The fog gateway, which is positioned in the middle of the Lab of the Intel Berkeley, uses the Internet to send data to the cloud. The EDaTAD approach uses the energy consumption model in Figure 6 [10]. The length of the data measurement is assumed to be 64 bits. Therefore, the size of the packet is the number of measurements that are needed to be transmitted to the fog node multiplied by the length of the data measurement.

The calculation of energy consumption is performed for sending a message consisting of L bits and a d distance as follow

$$E_{Transmit}(L,d) = (E_{elec} * L) + (\beta_{amp} * L * d^2). \tag{6}$$

The energy consumption for receiving L bits is determined as follows

$$E_{Receive}(L,d) = E_{elec} * L. \tag{7}$$

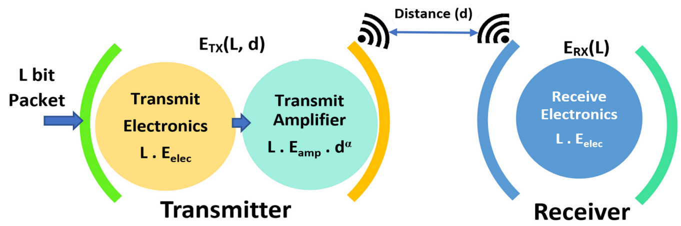


Figure 6: The First Order Radio Model.

Table 6: The parameters values of the simulation.

Parameter	Value
No. of nodes (M)	47 nodes
ρ	20, 50 and 100, 200, 500, 1000 readings
E_{elec}	50 nJ/bit
β_{amp}	100 pJ/bit/m^2
α	0.03, 0.05, 0.07, and 0.1
K	15

In this model, the power dissipation by the radio is E_elec to operate the transmitter or receiver circuits and β_{amp} for the amplifier of transmitter. The EDaTAD approach is compared to some existing techniques such as PFF [3], ATP [7], and Harb (FISHER, TUKEY, and BARTLETT) [8] approaches to ensure the improved performance of the EDaTAD approach. In this paper, the number of clusters (K) is selected according to the Elbow Curve Method. Table 6 introduces the parameters values of the simulation.

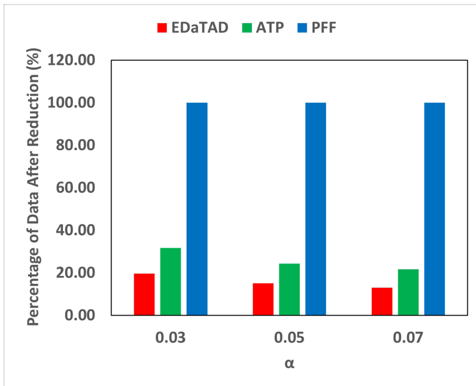
4.1 Percentage of Data After Applying Reduction Method

The principal mission of the sensor node is to gather the measurements and send them to the gateway of the next level of the network in an energy-saving way. The performance of the EDaTAD approach is investigated via decreasing the duplicating data at the sensor node. Figure 7 shows the percentage of data after applying the reduction method based on the LiReDaR algorithm of the EDaTAD approach.

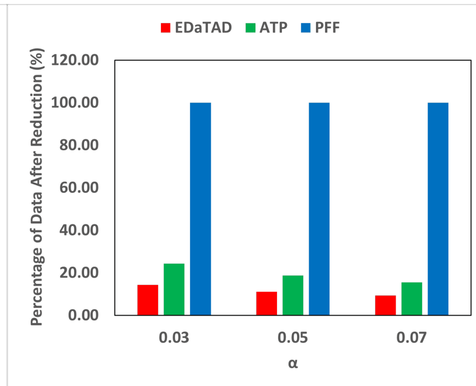
The EDaTAD approach decreases the unnecessary measurements before transferring them to the fog gateway from 38% up to 40% and from 80% up to 92% compared to ATP and PFF techniques, respectively. The findings of our LiReDaR algorithm of the EDaTAD approach indicate exhibits enhanced data reduction when the parameter α in the difference function grows, due to the increased correlation observed between the gathered data. Furthermore, when the number of data ρ increases, our approach significantly decreases the amount of gathered data.It has been demonstrated that the PFF method does not eliminate duplicate data at the sensor device before transferring it to the fog gateway. The results demonstrate the EDaTAD approach's effectiveness in eliminating repetitive data in order to save energy and enhance network performance.

4.2 Energy consumption of sensor device

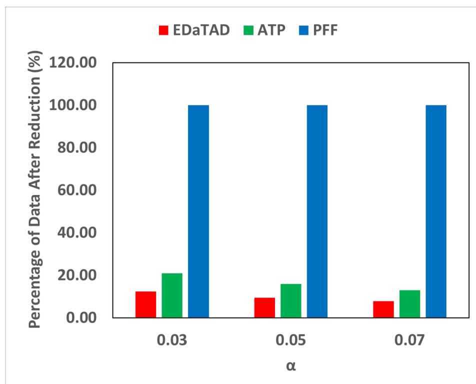
The energy consumption of sensor devices refers to a significant challenge that must be taken into account when developing a technique for a TI-based fog computing system. The EDaTAD approach investigates the consumed energy in this experiment. Figure 8 presents the consumed energy of the sensor device utilizing various data measurements sizes. It can be regarded that the presented EDaTAD approach reduces the sensor device's energy consumption from 84.60% up to 86.37% and from 87.23% up to 87.94% compared



(a) $\rho = 20$



(b) $\rho = 50$



(c) $\rho = 100$

Figure 7: Percentage of Data After applying Reduction method.

to the ATP and PFF methods, respectively. Accordingly, the sensor nodes using the EDaTAD approach save energy due to removing duplicated data, and this enables EDaTAD approach to outperform the other methods (see Figure 7).

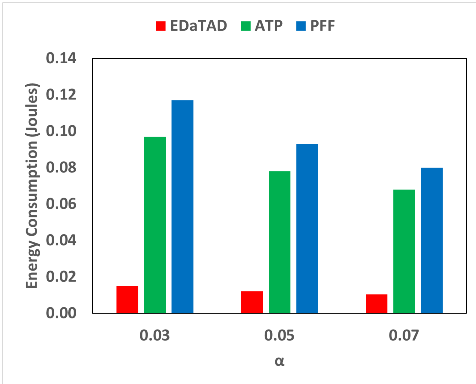
4.3 Data Loss Ratio

This section studies the impact of diminished data measurements on the data accuracy obtained at the fog gateway. In order to save energy, it is essential to reduce the size of transmitted data prior to actually sending it to the gateway, but it is also essential to sustain a sufficient rate of data accuracy at the gateway. The accuracy of data is indicated by the ratio of lost data that was collected at the gateway. It can be calculated as follow.

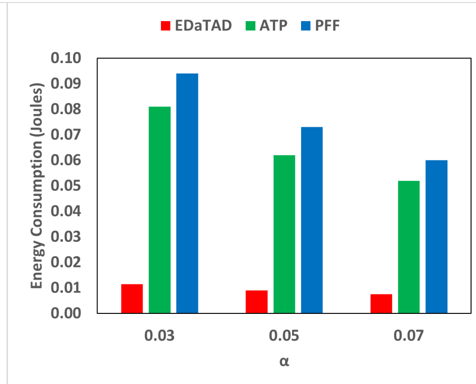
$$DataLossRatio(\%) = \frac{NLM}{TNM} * 100. \tag{8}$$

Where NLM and TNM refer to the number of lost measurements after applying data reduction and the total number of measurements before applying data reduction (size of original data), respectively. The data loss ratio is shown in Figure 9.

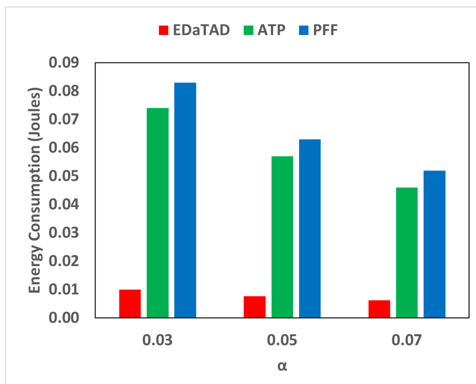
It can be noticed from the results in the Figure 9 that the EDaTAD approach decreases the size of lost data from 63.31% up to 91.21% and 52.83% up to 91.77% compared to ATP and PFF methods, respectively. Consequently, the EDaTAD approach reduce the size of transferred data and save energy while maintaining a sufficient ratio of accuracy.



(a) $\rho = 20$



(b) $\rho = 50$



(c) $\rho = 100$

Figure 8: Energy consumption of sensor device.

4.4 Transmitted Data Sets Ratio at Fog Gateway

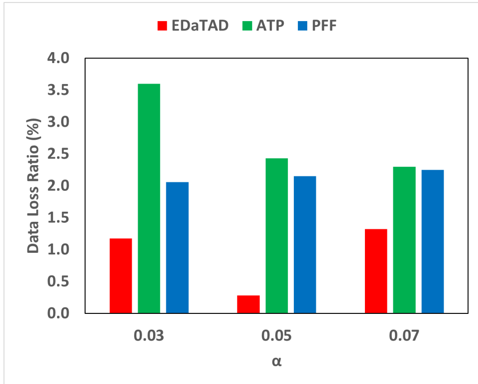
The similarities between the data reading sets received at the fog gateway can be significant. Therefore, it is necessary to eliminate the redundant data sets before transmitting the data to the cloud data center. This experiment explains the effectiveness of the proposed DaSeRE algorithm used by the proposed EDaTAD approach in removing these redundant data sets to save energy, reduce latency, and improve the performance of the network. Figure 10 shows the transmitted data reading set Ratio at Fog Gateway.

It can be noted from Figure 10 that the proposed EDaTAD approach decreased the data sets from 49.72% up to 62.91%, from 43.56% up to 60.91%, from 32.15% up to 41.19%, and from 20.51% up to 31.23% compared to PFF 0.8, PFF 0.75, Tukey, and Fisher methods respectively. It can be regarded that the Bartlett gives better results in only one case. However, the EDaTAD approach is better than the other techniques in general by removing the repetitive data readings sets and saving the energy at the gateway.

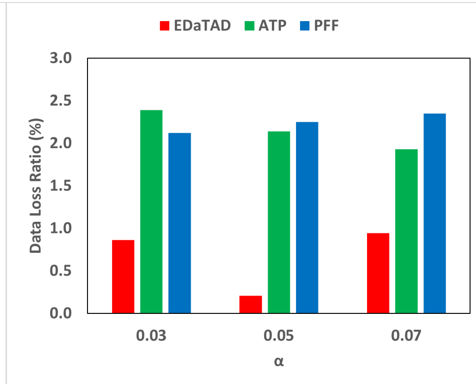
4.5 Consumed Power of Fog Gateway

Consumed power represents the most critical factor in most limited resources nodes. The performance of the proposed EDaTAD approach in terms of power consumption is studied and compared to some existing methods. Figure 11 shows the fog gateway consumed power .

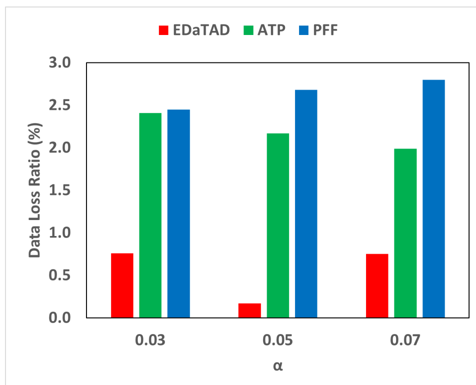
It can be noticed from Figure 11 that the proposed EDaTAD approach minimizes the energy consumption from 28.03% up to 74.23%, from 23.55% up to 73.36%, from 14.95% up to 58.86%, from 11.24% up to 61.90%, and from 1.37% up to 58.02% compared to PFF 0.8, PFF 0.75, Bartlett, Tukey, and Fisher methods respectively. The results show that the EDaTAD introduces better results than the other methods in saving energy at the fog gateway by removing redundant data readings sets after receiving them from the sensor



(a) $\rho = 20$



(b) $\rho = 50$



(c) $\rho = 100$

Figure 9: Data Loss Ratio.

devices.

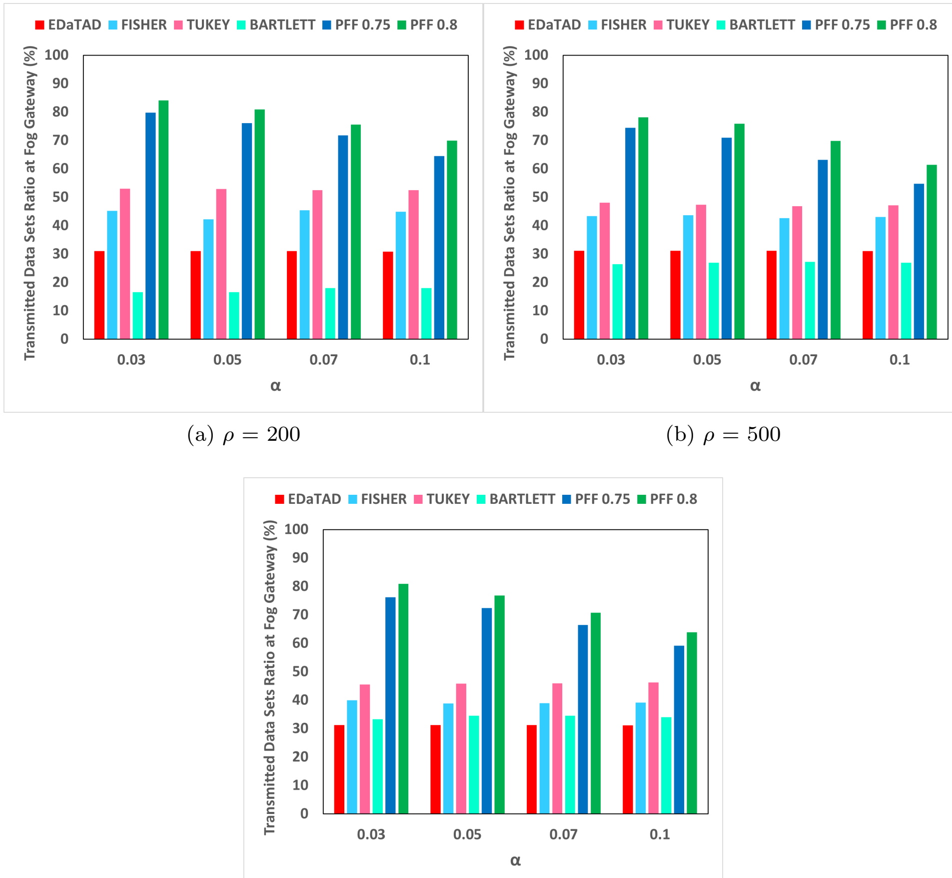
4.6 Decision-making

In this section, the proposed decision-making model that will be executed by the EDaTAD approach at the fog gateway is evaluated based on the received data at the fog gateway. In this experiment, it was considered the ρ and α are 20 and 0.05 respectively. Figure 12 refers to the aggregated scores that were calculated by our proposed EDaTAD approach using the decision-making model and the original data with only the decision-making model.

It can be seen from Figure 12 that the EDaTAD introduced aggregated scores for the collected data at the fog gateway close to the aggregated scores of original data in spite of a large reduction of data at the sensor level. The mean square error (MSE) between the aggregated scores of the proposed EDaTAD approach using the decision-making model and the original data with only the decision-making model is calculated as follows.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (ASx_i - ASy_i)^2 \tag{9}$$

Where ASx, ASy, and n refer to the aggregated score of proposed EDaTAD approach using the decision-making model, aggregated score of the original data with only the decision-making model, and the number of periods respectively. The calculated Mean Squared Error for the proposed EDaTAD approach is 0.087 that indicate close to 0. Therefore, EDaTAD approach achieves high reduction at the sensor nodes while the MSE for the aggregated score at the Fog gateway close to 0. This means the data loss is considered to be



(a) $\rho = 200$

(b) $\rho = 500$

(c) $\rho = 1000$

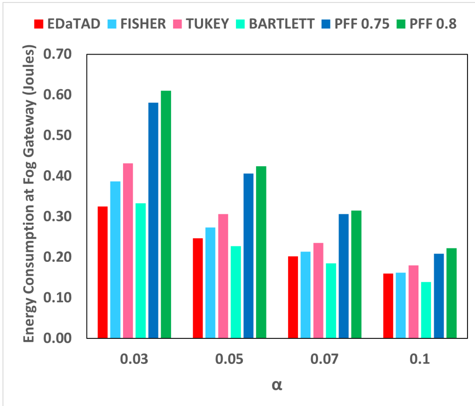
Figure 10: Transmitted Data Sets Ratio at Fog Gateway.

insignificant. Figures 13 (a) and 13 (b) show the decisions, which are taken based on the aggregated scores for both proposed EDaTAD approach using the decision-making model and the original data with only the decision-making model. These decision are made based on Table 4 using the calculated aggregated score.

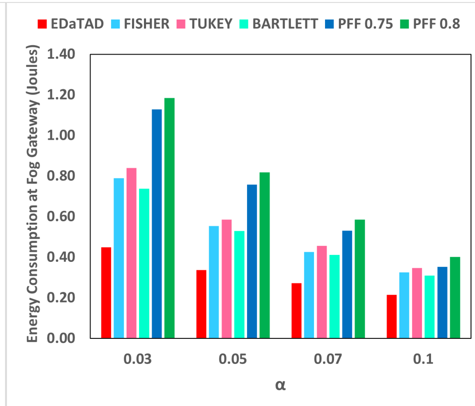
In Figures 13 (a) and 13 (b), the decision type for periods from 1 to 12 is 4 according to Table 4 because they have aggregate scores equal to 4. In period 194, the decision type is 2 because the aggregate score is within 1-2.99. The periods 236-257 and 292-298 have aggregate scores within 1-2.99. Therefore, they introduced a decision with 2. The periods from 299 to 3276 take the decision type 3 because they have aggregate scores within 3-3.99 according to Table 4. It can be observed from the Figures 13 (a) and 13 (b) that the high data reduction achieved by EDaTAD approach at the sensor nodes has no effect on the how the decisions are made at the fog gateway using the proposed decision-making model. The results show that the EDaTAD approach makes the same decisions when using original data without reduction. This ensures that the proposed EDaTAD approach could reduce the volume of data at the sensor nodes, save energy, and make suitable decisions at the fog gateway using the proposed decision-making model.

4.7 Latency time

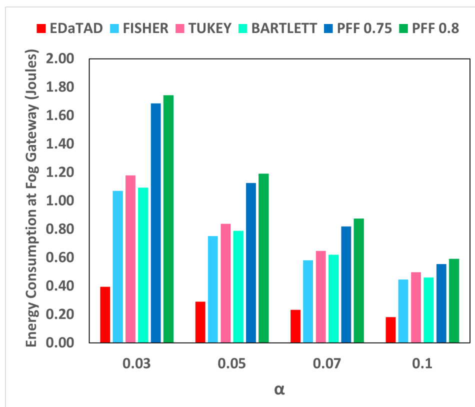
In this study, the delay time for the sensed data to move from the sensor node to the fog gateway is referred to as latency time. The sensed data is a set of data packets from the same time period. The major goal of lowering latency is to give the smart building staff a quick decision response regarding the state of the building. The latency is inversely proportional to the amount of sensed data. As a result, latencies are higher for bigger sensed data volumes. The latency has important impact on the response time at the Fog



(a) $\rho = 200$



(b) $\rho = 500$



(c) $\rho = 1000$

Figure 11: Energy Consumption at Fog Gateway.

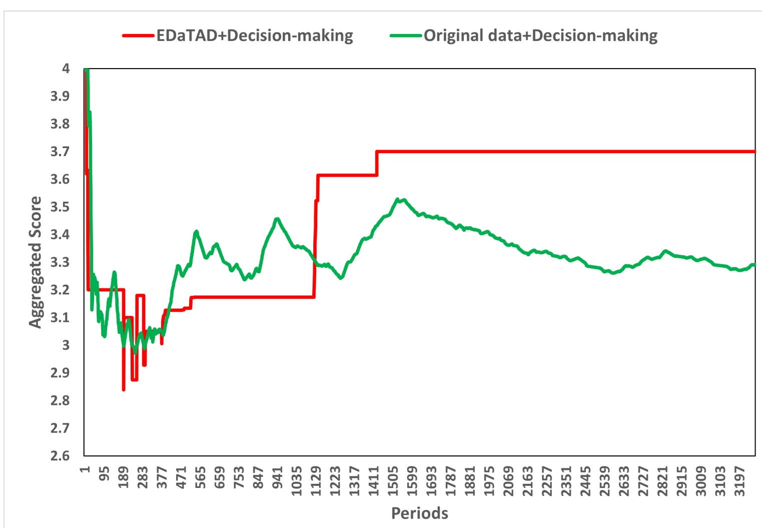
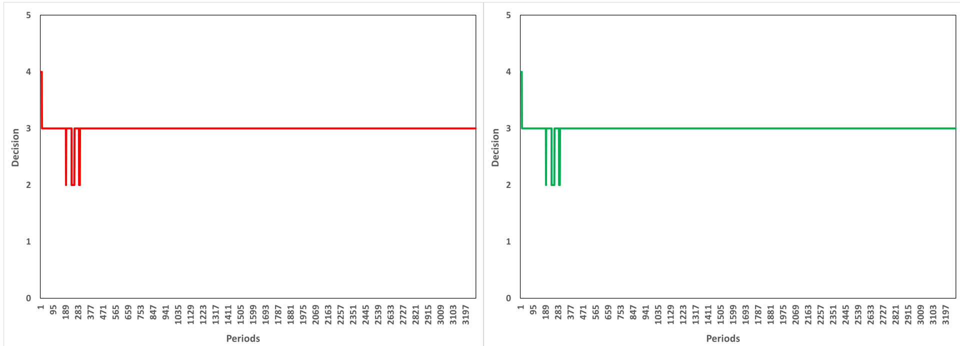


Figure 12: Aggregated score.

gateway. The latency time can be calculated using the following formula:



(a) a (b) b

Figure 13: The decisions. (a) Decisions by EDaTAD approach using the decision-making model. (b) Decisions by original data with only the decision-making model.

$$LatencyTime = t_{TX}^{Sensor} + t_{Link} + t_{Qu}^{Fog} + t_{Processing}^{Fog} \tag{10}$$

The t_{TX}^{Sensor} refers to the required time to send the sensed data during one period inside the sensor node. It is computed as follows:

$$t_{TX}^{Sensor} = \sum_{k=1}^{\kappa} \frac{\xi}{\Upsilon} \tag{11}$$

where κ, ξ, and Υ are the number of packets per period, the packet size, and transmission data rate respectively. The required time to send the sensed data to the Fog gateway through the transmission link is referred to as t_{Link} and it is calculated as follows.

$$t_{Link} = \sum_{k=1}^{\kappa} \tau \tag{12}$$

where τ is the required time to send one packet to the Fog gateway through the transmission link. The waiting time at the queue of the Fog gateway for each received data packet is referred to as t_{Qu}^{Fog}. The computation time at the Fog gateway is referred to as $t_{Processing}^{Fog}$. In this simulation experiment, we set the parameters ξ, Υ, τ, and t_{Qu}^{Fog} to 128 Bytes, 250 Kbps, 0.05 second, and 0.001 second respectively. For the sake of simplicity and during this simulation, we consider only one hop to the Fog gateway, and there is no packet loss. Figure 14 refers to the latency time during implementing the proposed EDaTAD approach.

In Figure 14, we computed the latency for two approaches: $With\_Using\_EDaTAD$ and $Without\_Using\_EDaTAD$ for different data sizes per period ρ. In the former, we compute the latency during implementing the proposed EDaTAD, while in the latter, the sensor nodes transmit all the sensed data to the fog gateway without reduction (original data). It can be observed from the findings in Figure 14 that the proposed EDaTAD lowers the time of latency between the sensor nodes and the Fog gateway because it greatly reduces the sensed data at the sensor nodes by removing the data redundancy whilst preserving the suitable quality of arrived data at the Fog gateway. The latency time is decreased by the proposed EDaTAD approach from 87.5% up to 91% in comparison with $Without\_Using\_EDaTAD$ for different sizes of data. This indicates the efficiency of the proposed EDaTAD approach in providing fast decisions to the smart building staff in the case of risks concerning buildings monitored by sensing devices.

5 Conclusion and perspectives

This paper suggests an Energy-aware Data Transmission Approach with Decision-making (EDaTAD) for Fog Computing-based IoT applications. The EDaTAD works on two levels in the Fog Computing-based IoT

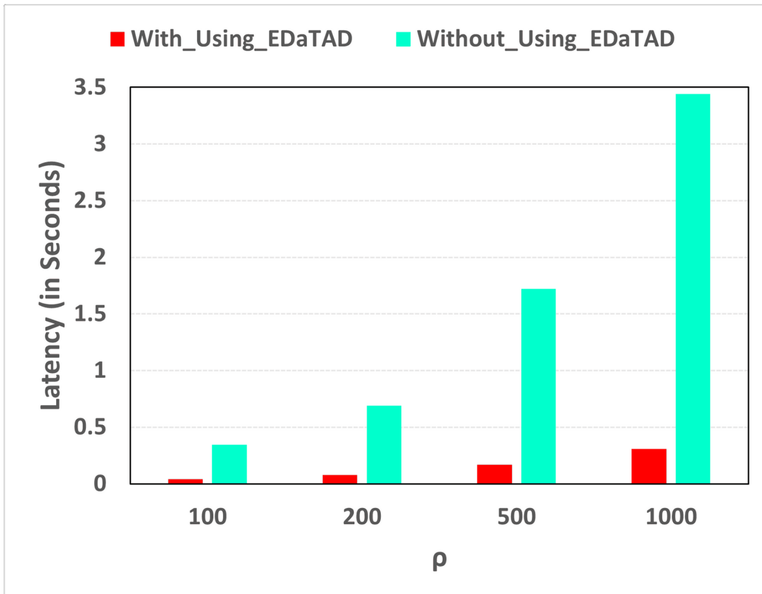


Figure 14: The Latency time.

architecture: sensor devices and fog gateway. The EDaTAD approach implements the LiReDaR approach at the sensor nodes to get rid of the repetitive data , while it executes DaSeRE algorithm at the fog gateway to eliminate repetitive data sets and evaluate the risk level of building environment to provide an accurate decision remotely for the monitoring staff. Several experiments have been performed to demonstrate the efficiency of the suggested EDaTAD approach. It can be noticed from the results that the EDaTAD improves the performance of the network and outperforms other methods in terms of data reduction ratio at sensor devices, energy consumption at sensor devices, data loss ratio, transmitted data sets ratio at the fog gateway, and consumed energy at fog gateway. Moreover, the results show that the EDaTAD approach determines the risk level of the building environment efficiently and can provide an appropriate decision for the monitoring staff remotely. The latency time is decreased by the proposed EDaTAD approach from 87.5% up to 91% for different sizes of data. In the future, we plan to implement lossy compression approaches to highly reduce the transmitted data and then implement deep learning techniques to enhance the quality of the reconstructed data in the fog or cloud layers.

References

[1] Shubhani Aggarwal and Neeraj Kumar. Fog computing for 5g-enabled tactile internet: Research issues, challenges, and future research directions. *Mobile Networks and Applications*, pages 1–28, 2019.

[2] Joseph Azar, Abdallah Makhoul, Rony Darazi, Jacques Demerjian, and Raphael Couturier. On the performance of resource-aware compression techniques for vital signs data in wireless body sensor networks. In *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, pages 1–6. IEEE, 2018.

[3] Jacques M Bahi, Abdallah Makhoul, and Maguy Medlej. A two tiers data aggregation scheme for periodic sensor networks. *Adhoc & Sensor Wireless Networks*, 21(1), 2014.

[4] CAA-QUEBEC. *Your Healthy-Home Guide. Available at: https://www.caaquebec.com/en/at-home/guides/your-healthy-home-guide/temperature-and-humidity-variations/*. 2021.

[5] Ramez Moh Elaskary, Mohamed Saeed, Tawfik Ismail, Hassan Mostafa, and Salam Gabran. Hybrid dct/quantized huffman compression for electroencephalography data. In *2017 Japan-Africa conference on electronics, communications and computers (JAC-ECC)*, pages 111–114. IEEE, 2017.

[6] H Harb, H Baalbaki, C Abou Jaoude, and A Jaber. Orchestration-based mechanism for sampling adaptation in sensing-based applications. *IET Smart Cities*, 2021.

[7] Hassan Harb, Abdallah Makhoul, Raphaël Couturier, and Maguy Medlej. Atp: An aggregation and transmission protocol for conserving energy in periodic sensor networks. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2015 IEEE 24th International Conference on*, pages 134–139. IEEE, 2015.

[8] Hassan Harb, Abdallah Makhoul, David Laiymani, Oussama Bazzi, and Ali Jaber. An analysis of variance-based methods for data aggregation in periodic sensor networks. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXII*, pages 165–183. Springer, 2015.

[9] Hassan Harb, Abdallah Makhoul, Rami Tawil, and Ali Jaber. A suffix-based enhanced technique for data aggregation in periodic sensor networks. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, pages 494–499. IEEE, 2014.

[10] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on*, pages 10–pp. IEEE, 2000.

[11] Marwa Ibrahim, Hassan Harb, Ali Mansour, Abbass Nasser, and Christophe Osswald. All-in-one: Toward hybrid data collection and energy saving mechanism in sensing-based iot applications. *Peer-to-Peer Networking and Applications*, 14(3):1154–1173, 2021.

[12] Ali Kadhum Idrees and Ali Kadhum M Al-Qurabat. Energy-efficient data transmission and aggregation protocol in periodic sensor networks based fog computing. *Journal of Network and Systems Management*, 29(1):1–24, 2021.

[13] Ali Kadhum Idrees, Rafal Alhussaini, and Mahdi Abed Salman. Energy-efficient two-layer data transmission reduction protocol in periodic sensor networks of iots. *Personal and Ubiquitous Computing*, pages 1–20, 2020.

[14] Ali Kadhum Idrees, Hassan Harb, Ali Jaber, Oussama Zahwe, and Mohamad Abou Taam. Adaptive distributed energy-saving data gathering technique for wireless sensor networks. In *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 55–62. IEEE, 2017.

[15] Sara Kadhum Idrees and Ali Kadhum Idrees. New fog computing enabled lossless eeg data compression scheme in iot networks. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–14, 2021.

[16] Alaa Shawqi Jaber and Ali Kadhum Idrees. Energy-saving multisensor data sampling and fusion with decision-making for monitoring health risk using wbsns. *Software: Practice and Experience*, 51(2):271–293, 2021.

[17] Ahmad Karaki, Abbass Nasser, Chady Abou Jaoude, and Hassan Harb. An adaptive sampling technique for massive data collection in distributed sensor networks. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 1255–1260. IEEE, 2019.

[18] Wazir Zada Khan, MH Rehman, Hussein Mohammed Zangoti, Muhammad Khalil Afzal, Nasrullah Armi, and Khaled Salah. Industrial internet of things: Recent advances, enabling technologies and open challenges. *Computers & Electrical Engineering*, 81:106522, 2020.

[19] Samuel Madden. Intel berkeley research lab. *http://db.csail.mit.edu/labdata/labdata.html*, 2004.

[20] James Newling and François Fleuret. Nested mini-batch k-means. *Advances in neural information processing systems*, 29:1352–1360, 2016.

[21] Alan Pears. *Strategic study of household energy and greenhouse issues*. Sustainable Solutions Australia, 1998.

[22] P Rajasekar and M Pushpalatha. Huffman quantization approach for optimized eeg signal compression with transformation technique. *Soft Computing*, 24(19):14545–14559, 2020.

[23] David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178, 2010.

[24] Alaa Shawqi Jaber and Ali Kadhum Idrees. Adaptive rate energy-saving data collecting technique for health monitoring in wireless body sensor networks. *International Journal of Communication Systems*, 33(17):e4589, 2020.

[25] ASHRAE Standard. 160 (2016). *Criteria for Moisture-Control Design Analysis in Buildings*, 1992.

[26] Cambrian T. National early warning score (news) 2: Standardising the assessment of acute-illness severity in the nhs, https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2, 2012. accessed: 2022-05-09. The Royal College of Physicians, 2017.

[27] Gaby Bou Tayeh, Abdallah Makhoul, Charith Perera, and Jacques Demerjian. A spatial-temporal correlation approach for data reduction in cluster-based sensor networks. *IEEE Access*, 7:50669–50680, 2019.

[28] Sotiris Vardoulakis, Chrysanthi Dimitroulopoulou, John Thornes, Ka-Man Lai, Jonathon Taylor, Isabella Myers, Clare Heaviside, Anna Mavrogianni, Clive Shrubsole, Zaid Chalabi, et al. Impact of climate change on the domestic indoor environment and associated health risks in the uk. *Environment international*, 85:299–313, 2015.

[29] Andras Varga. nesc language manual. *https://github.com/tinyos/nesc/blob/master/doc/-ref.pdf?raw=true*, 2003.

[30] Tian Wang, Jiandian Zeng, Yongxuan Lai, Yiqiao Cai, Hui Tian, Yonghong Chen, and Baowei Wang. Data collection from wsns to the cloud based on mobile fog elements. *Future Generation Computer Systems*, 2017.

[31] Peder Wolkoff and Søren K Kjærgaard. The dichotomy of relative humidity on indoor air quality. *Environment international*, 33(6):850–857, 2007.