International Journal of Wireless and Ad Hoc Communication

♦ **ISSN:** 2692-4048 **Vol:** 08, No .02, 2024 ♦ PP. 01–10
Article DOI: https://doi.org/10.54216/IJWAC.080201

FULL LENGTH ARTICLE

Enhancing Wireless Sensor Network Lifetime through Energy-Efficient Data Clustering and Compressed Forwarding in video Processing

Jayasudha A. R.[1] **Ramya S.** [2] **Vairaprakash S.** [3] **N. Kannaiya Raja**[4]
1 Department of computer applications, Hindusthan college of engineering and technology
2 Department of ECE, Sri Krishna College of Technology, India.
3 Department of Electronics and Communication Engineering, Ramco Institute of Technology, India.
4 Department of Computer Science, Ambo University, Ambo, Ethiopia
Emails: jayasudhacbe71@gmail.com ; ramyaprasadphd@gmail.com ; vairaprakashklu@gmail.com; kannaiya.raja@ambou.edu.et

Received: September 07, 2023 Revised: December 12, 2023 Accepted: March 01, 2024 ⋆ Corresponding author

ABSTRACT

Wireless Sensor Networks (WSN) play a crucial role in diverse data gathering applications, but face a significant challenge in the form of limited energy reserves within sensor nodes. Enhancing the network's Quality of Service, particularly its lifetime, is paramount. Prolonging the network's operational span hinges on mitigating energy consumption, with communication accounting for a substantial portion of nodal power usage. By reducing data transmission, not only can energy consumption be curtailed, but also bandwidth requirements and network congestion can be minimized. In the context of Wireless Sensor Networks, the Distributed Similarity-based Clustering and Compressed Forwarding (DSCCF) approach strives to construct data-similar iso-clusters with minimal communication overhead. This technique involves extracting trend and magnitude components from lengthy data series using an LMS filter, resulting in what is termed "data projection." Data similarity between nodes is assessed by measuring the Euclidean distance between these data projections, thereby facilitating efficient and low-overhead iso-cluster formation. To further economize intra-cluster communication, an adaptive-nLMS-based dual prediction framework is employed. During each data collection round, the cluster head holds instantaneous data for each cluster member, using either prediction or direct data communication. Furthermore, inter-cluster data is reduced via a multi-level lossless compressive forwarding technique. Impressively, this proposed approach has achieved an 80% reduction in data while maintaining optimal data accuracy for the collected information. The transmission of inter-cluster data exclusively occurs through a network backbone comprised solely of cluster heads. Initially, the cluster heads establish this network backbone. Each cluster head dispatches a link request query towards the sink through the backbone, receiving a link reply message containing path length and the weakest link of the path. The cluster head repeats this process for each available path, subsequently selecting the most optimal path based on the acquired information and its reliability in terms of link quality

Keywords: Wireless Sensor Networks (WSN) ▪ Energy Efficiency ▪ Data Clustering ▪ Data Projection ▪ Communication Overhead ▪ Network Lifetime Extension ▪ Machine Learning.

1. INTRODUCTION

Networks based on Wireless Sensor Data (WSN) are very scalable and adaptable. WSN is the preferred option for dispersed data collection because to its low upfront and ongoing expenses, ease of installation, and compatibility with mobile devices. WSNs are used in a wide variety of commercial [1] settings. In recent years, developments in VLSI design and RF technologies have positioned WSNs as a high-potential player in the fields of precision agriculture [2], intelligent transportation [3] and industrial remote sensing. The tremendous scalability and adaptability of Wireless Sensor Networks (WSN) have made them the preferred alternative to conventional networks. Since WSN requires less wiring and less upkeep, it is more cost-effective. Due to its low cost of installation and compatibility with mobile devices, WSN is rapidly becoming the technology of choice for widespread data collection.

Limitations in computer power and communication range are only two of the many difficulties inherent in putting WSN into practise because to its constrained energy supply. Nodes have a restricted lifespan because their power source, often a battery, has a limited energy budget. Since WSNs are both small and mobile-friendly, they are favoured for a broad range of uses. As a result, having a massively enormous Furthermore, nodes may be installed in a hostile or impractical area, making it hard or unpleasant to recharge the battery.

For WSNs to last as long as possible when used for large-scale data collecting, it is crucial that they use as little energy as possible while yet providing the level of detail and timeliness demanded by their applications. A WSN is made up of a network of autonomous sensor nodes spread out throughout an area, each of which sends data on one or more physical characteristics back to a central hub. Each sensor node in a WSN performs sensing, processing, and communication functions. The environmental node's wireless front end is a sensor or transducer. The data has been preprocessed and the computer has discharged its electrical output. It's possible that the prethen an RF transceiver sends to the base station is a straight k.

Due to its restricted communication and computation capabilities, WSN implementation presents several difficulties. Nodes have a restricted lifespan because their power source, often a battery, has a limited energy budget. Due to its portability and versatility, WSNs are increasingly being used in a broad range of settings. As a result, having a massively enormous Furthermore, nodes may be put in a hostile or unpleasant environment, making it either impossible or inconvenient to recharge the battery.

The energy consumption at high quality and resolution of the obtained data is of utmost importance when WSNs [4] are deployed for large-scale data gathering to allow meaningful analysis. On the one hand, the sensor network must accommodate a wide variety of uses. However, WSN also has significant challenges related to data accuracy and latency. Accuracy and low data latency are prioritised, so the node's remaining energy is conserved. The energy cost of distributed monitoring is strongly affected by the quantity and complexity of sensor node transmissions. The energy requirements of WSN are highest for wireless transmission and lowest for sensing. Contrarily, computing requires very little energy.

It has been noted that the cost of communication is substantially greater than the cost of calculation. In today's IoT world, devices [5] (are providing the ever-hungry internet with still more data) With a periodic data collection strategy, densely dispersed nodes monitor their surroundings and report back relevant information at frequent but brief intervals. This makes it possible to collect massive amounts of data. However, the substantial communication costs incurred by the WSN during its ongoing data collecting are the price paid for such complexity.

Reporting data at regular intervals is a significant drain on resources and, in a wireless network with limited capacity, leads to an increase in data collisions. data gathered for useful analysis. On the other hand, the sensor network's lifespan should be sufficient to meet the needs of the application. However, WSN also has significant challenges related to data accuracy and latency. As a result, the data collection procedure must use as little of the node's remaining energy as possible while yet achieving high quality results with as little delay as possible. The energy cost of distributed monitoring is strongly influenced by the volume of data broadcast from sensor nodes and the number of active sensor nodes at any one moment.

The energy requirements of WSN are highest for wireless transmission and lowest for sensing. However, computing requires a lot of power. One bit's worth of energy in MICAZ [6] mote requires 600nJ and 670nJ to transmit and receive, respectively, whereas computation energy per clock cycle is just 3.5nJ. Transmission uses 720nJ, reception uses 810nJ, and computing uses 840nJ of energy in TELOSB mote [7].

It has been noted that the cost of communication is substantially greater than the cost of calculation. As part of the Internet of Things [8] sensor nodes are now providing information to information-hungry cloud servers. In a periodic data collection strategy [9], densely placed nodes detect the environment and transmit the data of at short regular time intervals; only in this way is it possible to gather the finest data granularity. This paves the way for extremely composite examination of the data that has been gathered so far. However, this level of complexity comes at the expense of a decreased lifespan due to the significant communication costs involved during continuous data collecting. The lifespan of WSN is drastically shortened due to the periodic reporting of the observed data. In a wireless network with a limited amount of available bandwidth, an excessive amount of transmission leads to a decrease in throughput. Three is also a crucial factor. In time for the duration necessary to complete the application process. However, the process of data collection must be carried out with the highest possible precision and the least possible delay in mind. The energy requirements of a distributed monitoring system are very sensitive to the volume of data being relayed in real time by active sensor nodes.

The energy requirements of WSN are highest for wireless transmission and lowest for sensing. Computing, on the other hand, is the energy requirements of tion are 720 nJ, 810 nJ, and accordingly in MICAZ [10]. Densely deployed nodes detect the environment and transmit the data of, only by which a narrowest data granularity composite data analysis on the observed data (despite the greater communication cost) Korpeolu2003. However, the significant expense of communication required to attain this level of complexity shortens the lifespan of the lifespan of WSN is drastically shortened because to the large amount of power it requires to regularly report the observed data.

Excessive communication caused by spatial and temporal similar data in a wireless network with limited capacity. Energy conservation with lowest tolerance under the first rule of geography, which states that "everything is connected to everything else, but close things are more related than distant things," may be achieved by making use of this characteristic of WSN data. This statistical finding suggests that data similarity between sensor nodes (data similarity between t spatial correlation) improves as their physical distance from one another decreases. Due to spatial correlation, data transfers between nearby nodes are sometimes redundant. The similarity between two or more successive measurements of node time is quantified by a metric called temporal correlation. A large quantity of redundant information is added to the node's data series because of the sensor data's temporal correlation. The nodes in a WSN are closely spaced and share information with a high degree of temporal and geographical correlation. High spatial and temporal correlations in WSN lead to a great deal of duplicate information being delivered [11]. These useless duplicates take up a lot of bandwidth and storage space on the network for no good reason. Thanks to Spatio's widespread presence, transmission may be aggressively lowered for energy conservation without significantly impacting the accuracy of observed data. The integrity of transmitted data may be maintained while still conserving a significant amount of node energy. Saving bandwidth and decreasing congestion are two additional benefits of eliminating this kind of duplicate or comparable material.

Sensory information has a natural tendency toward consistency over space and time. By taking use of this characteristic of WSN data, energy may be saved with just a small margin of error. "Everything is connected to everything else, but close things are more related than distant things," argues Tobler's first rule of geography [12]. Spatial correlation refers to the degree to which data similarity between nearby nodes over a given region rises as the distance between sensor nodes decreases [13]. Due to spatial correlation, data transfers between nearby nodes are sometimes redundant. The similarity between two or more successive measurements of node time is quantified by a metric called temporal correlation. A large quantity of redundant information is added to the node's data series because of the sensor data's temporal correlation. Densely distributed nodes in a WSN send data at fast rates, creating strong spatial and temporal correlations. High spatial and temporal correlations in WSN lead to a great deal of duplicate information being delivered. These duplicates contribute nothing to the knowledge base but take up a lot of valuable network space.

The widespread presence of Spatio-temporal correlation in the sample data enables aggressive data transmission reduction for energy savings without significantly degrading the quality of the observed data. Therefore, a large amount of nodal energy may be preserved by reducing duplicate information without compromising the accuracy of the data collected. Saving bandwidth and decreasing congestion are two additional benefits of eliminating this kind of duplicate or comparable material. Sensory information always has 4 dimensions. The precision of energy data collection may be improved by taking use of this feature of WSN data. Everything is connected, yet nearby items are more so than those far away, as stated by Tobler. This statistical finding suggests a rise in data correlation, As cited in Villas et al. Spatial correlation describes the relationships between nodes in each region. Due to spatial correlation, data transfers between nearby nodes are sometimes redundant. The degree of similarity between two node measurements at regular intervals in time is defined as the temporal correlation.

A large quantity of redundant information is added to the node's data series because of the sensor data's temporal correlation. Nodes in a WSN have high sampling rates, leading to accurate data. Because of strong geographical and temporal correlations in WSN, a large proportion of the overall sent data is redundant information with little in the way of informational value. Data transmission may be lowered aggressively for energy saving thanks to temporal correlation in the sample data, and a considerable amount of nodal energy can be preserved by avoiding duplicate data transmissions. Congestion and bandwidth use may be drastically decreased by eliminating unnecessary data. Improved data quality thanks to more frequent sampling. However, not all the data that was sampled should be sent. Due to the increased temporal correlation (values), substantial effort is spent on duplicate data when increasing the sampling frequency of a sensor node. Massive energy savings may be achieved by not sending any data that isn't essential. High temporal correlation allows for confident estimation of observations. It is possible to foretell a sensor node's future readings by analysing its recent past readings and the pattern of its readings. Using these procedures, sensor readings may be sent with a high degree of certainty. Estimation and prediction systems in the time domain work to lessen the burden on the system's power supply by replacing some of the most draining operations with others that use less juice. Denser placement of nodes enhances spatial correlation between nearby nodes, which may improve network reliability and coverage efficiency in WSN. This redundancy in space-time (numbers of nodes, quantities of energy, etc.) is especially useful when the nodes are physically near together. As a result, information about a sensor node may be reliably inferred from that of its immediate neighbours. Sensors tend to have a high degree of spatial correlation. Their strong geographical connection is supported by the size and trend similarity of the data g. Only a small fraction of the sensors in this clustered collection need to provide values for the group to be accurate.

Related Work

WSNs are used in a wide variety of commercial and industrial settings [14] applications in environmental sensing and biomedical imaging. Smaller wireless sensor nodes are now possible thanks to developments in microelectronics and wireless technology [15] and these nodes are being deployed in a wide variety of environments to track and record changes in physical quantities like temperature, pressure, vibration, light intensity, and sound volume. WSNs have emerged as a high potential player in the fields of precision agriculture

[16] cognitive environments [17] military surveillance, and industrial robotics as a result of recent advances in VLSI design and RF technologies. thus, wireless sensor Designing energy-efficient features into wireless sensor networks is an important challenge. In general, various works have attempted to maximise the lifespan of WSNs. The size of the power source is kept minimal since WSNs are favoured in many applications because of their compact design and mobility compatibility. Nodes in a WSN cannot have high-performance CPUs or long-range radio transceivers due to the network's limited energy supply. The network's longevity is also impacted by the size of the power source. Efficient use of on-board energy is crucial for extending lifespan. Numerous efforts have been done on energy saving in WSN due to the fact that limited energy is a barrier for most WSN applications.

The percentage of time that nodes are really active throughout a data gathering cycle is known as their duty cycle. The network's energy consumption may be drastically cut down by lowering the duty cycle. Only when the mobile data collector is in close proximity to the sensor node will it send data. This technique reduces energy consumption by preventing the transmission of data across vast distances. Data from a WSN have many similarities with one another in terms of both location and time, therefore an approach that takes use of these similarities may help save energy without compromising data quality.

WSN implementation is difficult because of the energy constraints imposed by its limited resource. Nodes have a restricted lifespan because their power source, often a battery, has a limited energy budget. Since WSNs are both small and mobile-friendly, they are favoured for a broad range of uses. Having a massive generator is thus impracticable. Furthermore, nodes may be installed in a hostile or impractical location, making it hard or unpleasant to recharge and replace the battery. When WSNs are used for massive data collecting, it is crucial to minimise the number of failures to a minimal. The geographical and temporal resolution of an application is what drives the design of WSNs for persistent, decentralised data collection. Distributed sensor nodes (SDNs) form a WSN, and they all work together to provide data on one or more physical parameters back to a central server. Each sensor node in a WSN is an independent system capable of detecting, processing, and exchanging data with other nodes in the network. The wireless node's "front end" is a sensor or transducer that collects environmental data. The data has been pre-processed and the computer has discharged its electrical output. An RF transmitter then sent the information to the station's control room. Both direct and multi-hop networks may be used for data transmission between the sensor and the base station. The node components function like a battery in that they have a finite amount of juice.

There are three stages of Distributed Similarity based Clustering with Compressed Forwarding (DSCCF). Data similarity clusters are built in the first step. Using an LMS filter, each node builds its own data projections. Energy-aware delay has been proclaimed by the higher-level nodes. The non-calculate the Euclidean distance between their data projection and the CH's data projection and then link the two sets of data using the CHs that have the closest projection.

2. PROPOSED WORK

Continuous reporting of sensor readings to the sink node at predefined time intervals is the most typical mode of data gathering in operational WSNs. This technique is a time-driven data collecting system since the data is gathered and provided at regular intervals regardless of whether the data changes. Measurements and low-entropy data transmissions to the sink node are the main functions of most sensor networks. In the dual prediction approach, the sink node accurately predicts the sensor node's future readings using a time series prediction model, rather than interacting directly with the sensor node. Consequently, it may be possible to avoid the tremendous expense of maintaining many radio transmissions between the sink and the nodes. The strategy's principal objective is to transmit the chosen sample set. In this setup, every sensor node has its own prediction model trained with the data it has acquired.

With the DPF, there is much less communication between the sensor node and the sink node, and the collected data is guaranteed to be within the user-specified error limit. Nevertheless, the efficacy of the prediction model in matching the sensor-acquired time series is crucial to the communication advantages of the DPF. Given that the sink node has access to the most recent historical samples, the DPF's principal function is to estimate future sensor readings by executing the same prediction models at both the source and sink nodes. If the expected value differs from the real sensor data by more than a certain error threshold, a revised model is sent to the sink.

Nodes in the DPF's sink and source nodes both execute prediction models; the latter uses them to approximate the real sensor signal by a consistent amount.

If the predicted value is within the user-selected error threshold, no updates will be sent during prediction until there is a significant deviation from the sensor node's readings.

Sharing sensor readings occurs when the gap between the two becomes too large to ignore and continues until the prediction is in line with the goal value. In contrast to the status quo method of monitoring.

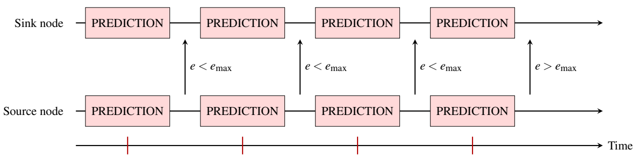


Figure 1. Dual prediction framework

The source node will not transmit data until the prediction departs from the target data by a greater amount than the maximum deviation threshold, emax. This means less expensive periodic radio broadcasts. You can see how the dual prediction framework works in Figure 1. In the DPF, the filters for the source node and the sink node are identical. There are three distinct modes of operation: initialization, normal, and stand-alone. Before entering operational or standalone mode, a node only runs the initialization mode once.

In the initialization mode, the source node regularly transfers the observed data to the sink at sampling instants t. A model for temporal prediction using an adaptive filter is simultaneously constructed by the source node. This mode's power consumption is as follows.

$Einit = Etx + Efx + Epred$Einit =Etx+Efx +Epred (1)

Dual prediction energy consumption is calculated as

$Esa = 2Epred$Esa =2Epred (2)

When the error exceeds emax, the mode will shift to When all systems are operational, information is transferred to the sink.

Normal functioning; Initial Configuration. With the use of weight adjustments, the source-based prediction engine brings the forecast closer to the target value. When convergence is reached in the prediction, the mode goes back to standalone. A typical calculation for electricity consumption is

$Enorm = Etx + Edx + Epred$Enorm =Etx+Edx+Epred (3)

Adaptive Filters

Everything that analyzes signals or sends data over a network relies on filters. A device or procedure may be used to filter out unwanted elements in a signal. To model the input-output relationship of a system, (b) decompose a signal into two or more sub-band signals for sub band signal processing, (c) modify the frequency spectrum of a signal, and (d) restrict the signal to a required frequency band or channel, such as in the case of an anti-aliasing filter.

The use of digital filters allows one to go more deeply into a time series. A digital filter may extract the trend, seasonality, and magnitude components of a time series by altering its coefficients until they converge with the data. An adaptive filter is a software tool that repeatedly represents the input-output signal relationship. Figure 2 shows an example of an adaptive filter, which uses an adaptive algorithm to automatically modify its filter coefficients. In many cases, real-time applications do not have access to characteristic information on the phenomenon being observed. An attractive option is the use of adaptive filters, which are self-adjusting systems that analyze data using recursive algorithms. In a self-regulating filter, the incoming reference signal acts as the regulator. The filter uses a training vector to interpret the desired response in numerous ways. A linear combination of the input signal components and the beginning filter coefficients is used to generate the initial value that the filter produces. An error signal may be generated and used to adjust the filter's settings by comparing the filtered output with the desired value. To achieve stability, the filter's coefficients must be fine-tuned continuously.

Every iteration, the filter inputs some time series historical data, convolves it with the weight coefficients, and then outputs a future value. As the prediction error approaches zero, the filter gradually reduces it by adjusting the value of its weights. The prediction filter is said to have converged when its output closely resembles the original data series. A filter's convergence speed is its learning rate for the time series. Now that the filter has all the time series features, it can use them to predict how the time series will go in the future. A better learning filter will have a smaller prediction error.

While statistical parameters evolve over time, time series and their corresponding noise remain relatively constant. The transfer function of an adaptive filter is adjusted by an adaptation algorithm that monitors the trend of the time series. It might be stable, or it could be in the process of converging. When operating in steady state, the wiener filter converges to the target signal. As the filter converges, it should adapt to new circumstances by modifying the filter coefficient to reflect the passage of time. The purpose of an adaptation procedure is to adjust the filter settings (transfer functions) so that they work better in a changing environment. One performance metric is the mean squared error.

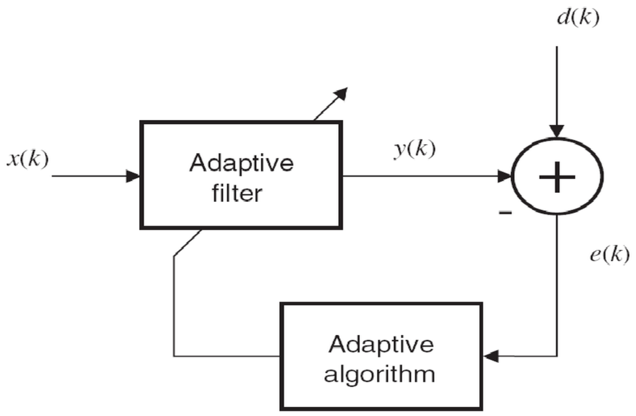


Figure 2. Adaptive Filter.

LMS Adaptive Filters

There is a plethora of techniques used for data prediction in WSN. Even though these techniques have reduced power consumption, they have only worked thus far since they needed a priori data to accurately predict future values. The proposed alternative relies on a model-free algorithm, which allows nodes to operate independently of one another and the model's global parameters.

An alternative approach to data reduction using the LMS adaptive algorithm is presented in this work. The LMS method employs two main operations to process the signal; it is an example of a linear adaptive algorithm. Every time the filter runs, its output is checked against the target signal to obtain a rough idea of the error. This rough estimate is then used to adjust the weight of the filter in the parametric model.

The difference between the input and output is computed using the LMS algorithm. At each iteration, LMS modifies the weight vector in the opposite direction as the output deviation to reduce the mean square error. An LMS adaptive filter is shown in Figure 3 as a block diagram. For LMS, there's no need to compute correlation functions or invert matrices. Thanks to this, the work becomes quite simple.

According to what is said, LMS is a stochastic gradient approach. It calculates the gradient of the error performance surface with respect to a vector that might vary randomly. Thanks to LMS's non-linear feedback, convergence may happen quickly with little computing cost.

The LMS adaptive approach is lightweight in terms of computation and memory use. Despite how simple it seems, the LMS algorithm gets excellent results. Moreover, in contrast to previous research, the LMS technique does not need any prior knowledge or modeling of the statistical characteristics of the observed signals. Consequently, this method may be used for a wide variety of real-world phenomena.

Nodes may function independently of a centralized authority

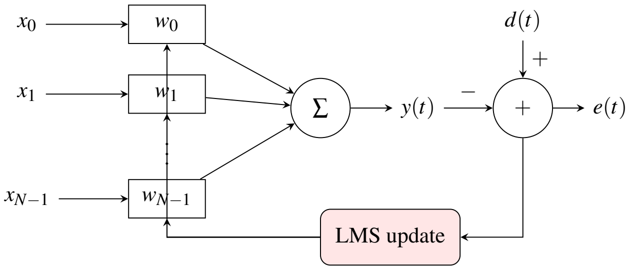


Figure 3. Block diagram of LMS adaptive filter

if they use the proposed strategy, which relies only on local data for prediction purposes. Thanks to its flexibility, the LMS filter may be integrated with many current data collection techniques. For example, it can be used in combination with in-network data aggregation tools.

Separate prediction models are kept by the sensor node and the sink node in the Dual Prediction Technique (DPF). We may write the prediction model of the sensor node as C-b and the prediction model of the sink node as C-Z. Estimates of future sensor data are made using these models.

At time tt, the sensor node predicts the future value of the sensor reading as $yt+1=Fsxt$yt+1=Fsxt, where xtxt is the historical data collected up to time tt. Similarly, the sink node predicts the future value as $yt+1=Ftxt-M,\dots,xt-1$yt+1=Ftxt-M,...,xt-1, where $xt-M,\dots,xt-1$xt-M,...,xt-1 are the most recent historical samples available at the sink node.

The prediction models aim to minimize the prediction error. Let's denote the actual sensor reading at time $t+1$t+1 as $yt+1$yt+1. The prediction error for the sensor node is given by $es=yt+1-yt+1$es=yt+1-yt+1, and for the sink node, it's $et=yt+1-yt+1$et=yt+1-yt+1.

The communication between the sensor node and the sink node occurs when the prediction error exceeds a certain threshold. Let's denote this threshold as εε. If $es>\varepsilon$es>ε, the sensor node sends an update to the sink node containing its latest prediction model. Similarly, if $et>\varepsilon$et>ε, the sink node requests an update from the sensor node.

The primary goal of the DPF is to minimize communication while ensuring that the prediction error remains within the user-specified error boundary. This is achieved by updating the prediction models only, when necessary, i.e., when the prediction error exceeds the threshold εε.

The communication overhead of the DPF depends on the accuracy of prediction models Fs and Ft. If the models accurately capture the underlying dynamics of the sensor readings, fewer updates are required, leading to lower communication overhead.

In comparison to the status quo monitoring technique, where data is continuously reported to the sink node at predetermined intervals regardless of changes in the data, the DPF reduces communication overhead by dynamically updating prediction models only when necessary. This results in more efficient utilization of network resources and reduced energy consumption, especially in large-scale operational WSNs.

3. EXPERIMENTAL ANALYSIS

The ASAL-nLMS process is modeled in MATLAB. There are two ways to measure the efficiency of dual predictions: the amount of data sent and the average difference between forecasts and observations. Other metrics used to measure performance include the number of model reconstructions and the average number of rounds to converge. The actual performance of the predicted work is evaluated using three criteria. The proposed system is first evaluated by comparing its performance with several temperature datasets that exhibit different levels of correlation (accessible at: http://www.db.csail.mit.edu).

The new system is tested using a variety of non-linear physical factors, such as battery voltage, humidity, light, and wind speed (http://www.ndbc.noaa.gov). We also look at the resource sharing capabilities of clustered networks. In this case, we use a LEACH-based clustering network to estimate how effective the proposed filters.

4. PERFORMANCE ANALYSIS ON TEMPERATURE DATA SET

The first 1000 data points from series 1, 2, 3, 4, and 5 of the real-world sensor node data from Intel-Berkley laboratories are utilised for assessment (Available:.http://www.db.csail.mit.edu). Table 1 displays the ASAL DPF simulation settings.

Table 1. Simulation parameters for ASAL DPF

Parameter	Value
N	20
rth	0.8
C	0.5
emax	0.25°C, 0.5°C, 0.75°C, 1°C, 1.25°C

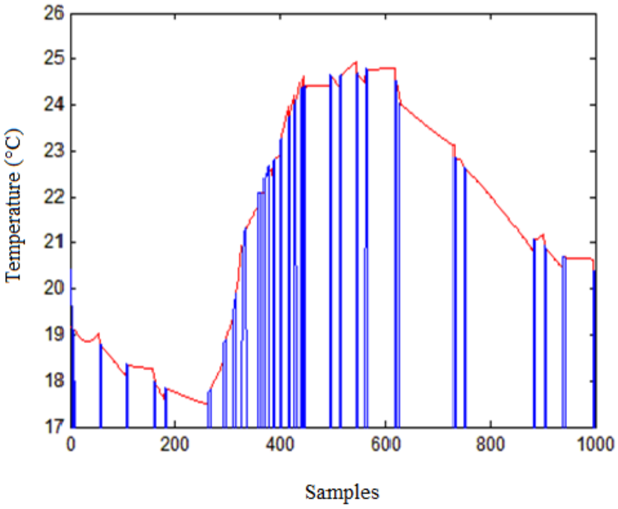


Figure 4. ASAL-nLMS transmission instances

The adoption of ASAL-nLMS has resulted in a halving of the needed transmissions, from 1000 to 90. Figure 4 depicts the data transmission process from one node to another. Examples of ASAL-nLMS transmissions that were recreated from the data within a tolerance of 0.25°C are shown in Figure 5. D in ASAL-nLMS may take on values between 2 and 20 when running normally. Adaptively modifying the step size reduces the overshoot at the convergence point. Over a 20-point sliding window, the data's linearity is assessed. If rth > 0.8, the greater filter length is used; otherwise, the shorter filter length is chosen. This change in length allows for more precise predictions across the linear and dynamic domains. The outcomes of comparing the actual data with the predictions of the ASAL-nLMS are shown in Figure 3.2..

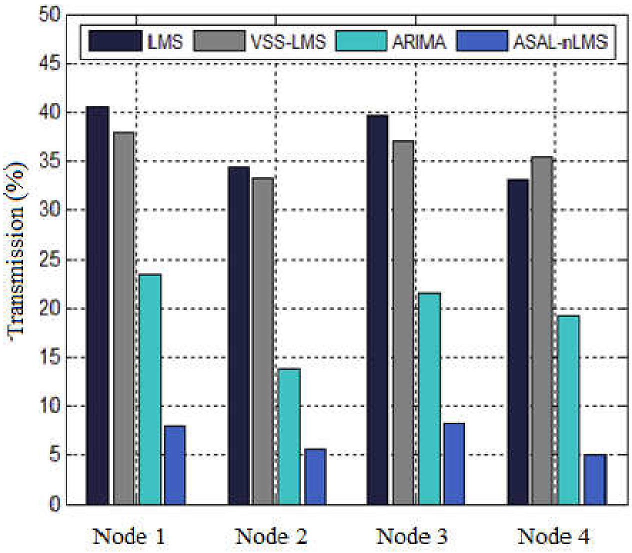


Figure 5. Data transmission comparison for different prediction methods

(emax=0. 5°C)

Table 2. ASAL-nLMS performance on battery voltage data set

emax (V)	RMSE (V)	% Transmission
0.05	0.01	50
0.1	0.02	30
0.15	0.035	20
0.2	0.85	28

Table 3. ASAL-nLMS performance on humidity data set

emax (%RH)	RMSE (%RH)	% Transmission
0.2	0.5	40
0.4	0.15	20
0.6	0.19	5
0.8	0.032	5

Table 4. ASAL-nLMS performance on light data set

emax (Lx)	RMSE (Lx)	% Transmission
10	2	45
20	5	35
30	8	20
40	10	20

Table 5. ASAL-nLMS performance on wind speed data set

emax (kmph)	RMSE (kmph)	% Transmission
0.2	0.18	65
0.4	0.1	55
0.6	0.2	40
0.8	0.25	35

Model reconstruction durations could be shortened due to the linearity of data sets like humidity and battery voltage. But further model reconstructions during the same time period are required due to the unpredictability of light and wind speed data sets. A bigger error margin allows for more leeway in the acceptance of deviations from the prediction. As a result, the rates of model rebuilding are considerably reduced. Reduced error tolerance leads to stricter limitations, which in turn causes more frequent model reconstructions and a substantial decrease in the mean prediction error. Recreating models takes a fraction of the time when data is collected in a linear fashion. Figure 6 illustrates how the linearity of the data set reduces as the level of randomness increases, as the result is over 100 model reconstructions for 1000 data points.

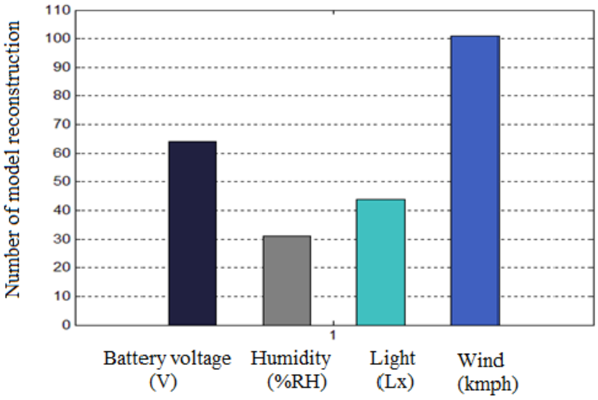


Figure 6. Quantity of ASAL-nLMS model reconstructions

Various ASAL-data sets of different sizes are shown in Figure 7 nLMS representing convergence periods. Because of its linearity and small standard deviations, humidity data could be more suited for fast model rebuilding. Reconstructing a model from scratch sometimes requires four iterations when working with fresh data sets. Model reconstruction converges at a slower rate when data contains outliers. We rebuild the model with few iterations if the data trend is inconsistent.

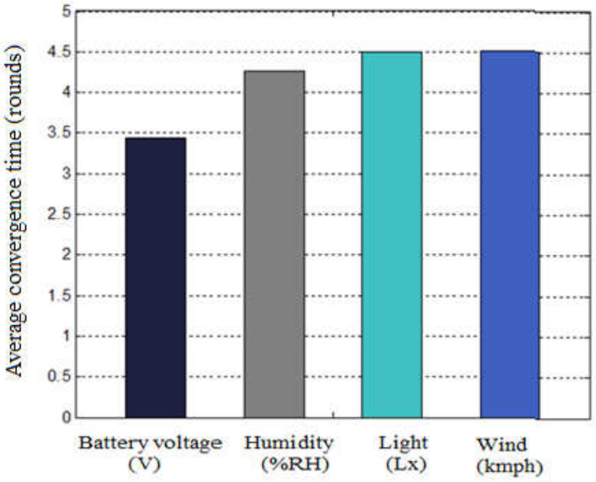
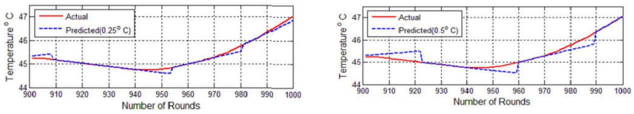


Figure 7. Average convergence time of ASAL-nLMS for different data sets

At an emaxof 0.25°C, Figure 10 shows the data transmission percentages for different nodes. In terms of data reduction,



(a) (b)

Figure 8. Original and predicted data of adaptive nLMS: (a) 0.25 °C, (b) 0.5°C

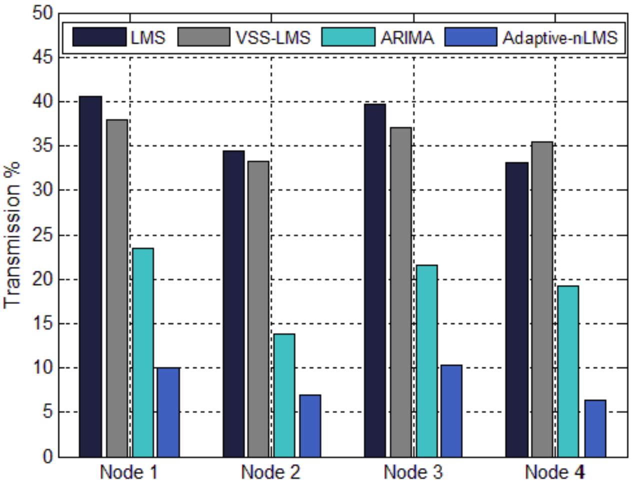


Figure 9. Data transmission comparison of different prediction approaches (emax=0. 25°C)

adaptive nLMS is about 20% more effective than LMS and VSS-LMS, and 50% more effective than ARIMA (Santini & Romer, 2006). Normalization and adjustment of step size are mostly responsible for this substantial drop in data. In a dynamic context, data reduction should decrease as a function of error threshold and data dynamics to maintain the needed degree of accuracy in the predicted signal.

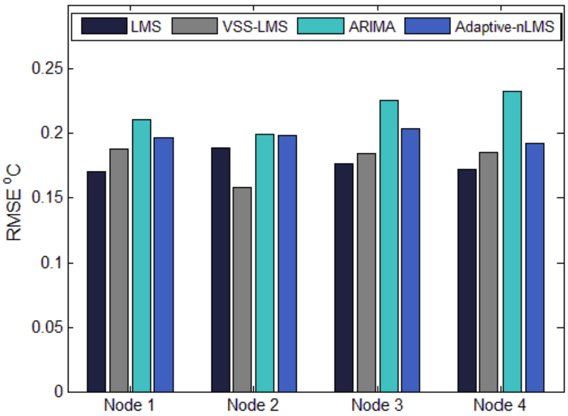


Figure 10. RMSE comparison of different prediction approaches (emax=0. 25°C)

In data aware clustering, testing the cluster's integrity is essential. The clusters are modified by combining or dividing them based on how comparable the data is. To avoid overwhelming CH nodes, all dynamic clustering methods need periodic cluster generation. Cluster setup and maintenance operations should consequently have little overhead. A data-aware clustering model's clustering performance may be roughly estimated by counting the control packets required for cluster establishment and maintenance. Counting the number of packets sent with different levels of network data correlation is one way to gauge clustering's efficacy.

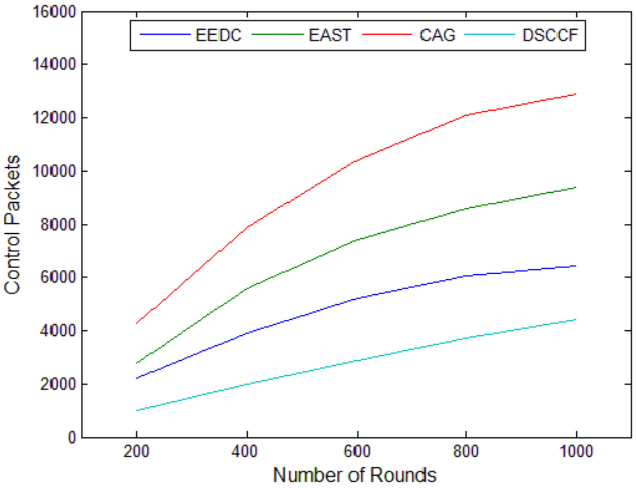


Figure 11. Comparison of clustering overhead for different clustering methods

This node might be a member, a representative, or a coordinator depending on the data it provides. A single "sink" node in an EEDC network receives massive data sets from all the nodes and uses them to assess how similar they are based on their trajectories over time. Due to the huge data dimension and the extended distance conveyed during each clustering cycle, the communication cost of being an EEDC is considerable.

Building data-similar clusters with little control message overhead is possible using the suggested approach. Each member of the cluster is only required to send one message (an affiliation) and the CH node only must send one message (an announcement) to construct the cluster according to the suggested manner. Because DSCCF checks the veracity of each cluster member's DCR automatically, cluster management is made easier. The proposed approach calls for broadcasting an equal number of clustering signals to the number of nodes in the network. Using DSCCF to build a network's backbone doesn't cost much more than building the original cluster. The control overhead is low since the backbone is built using just CHs.

Data compression performance is strongly affected by the cluster size and the degree of spatial correlation between sensor nodes. Our focus here is on the message cost per degree. Clusters of varying sizes and degrees of geographical closeness are shown.

As a cluster grows, the likelihood that its members will share characteristics increases. Because of the stronger spatial correlation, it is common for neighboring nodes to reflect changes in data at the same rate as the original node. Therefore, several data values may be condensed into one record using run length encoding. Consequently, compression efficiency is enhanced.

The suggested approach relies heavily on compression at the CH. Here, CH doesn't assume that cluster members share any kind of shared geographical data. Rather, it compresses data depending on how similar it is, using data that is indicative of the complete cluster. Consequently, this kind of compression is effective and dependable. While data is compressed without loss while moving between clusters, it is compressed with loss when collected inside a cluster. The space savings

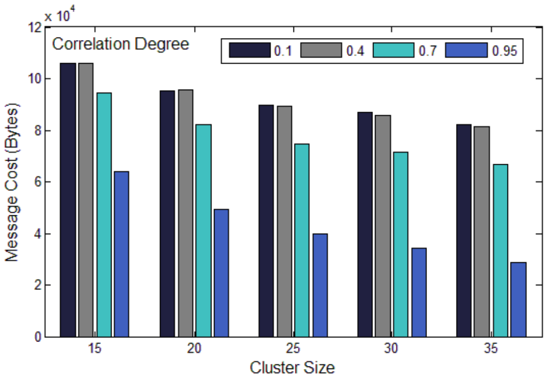


Figure 12. Cost of messages for various cluster sizes with different degrees of correlation

and accuracy benefits of lossless compression are mutually beneficial.

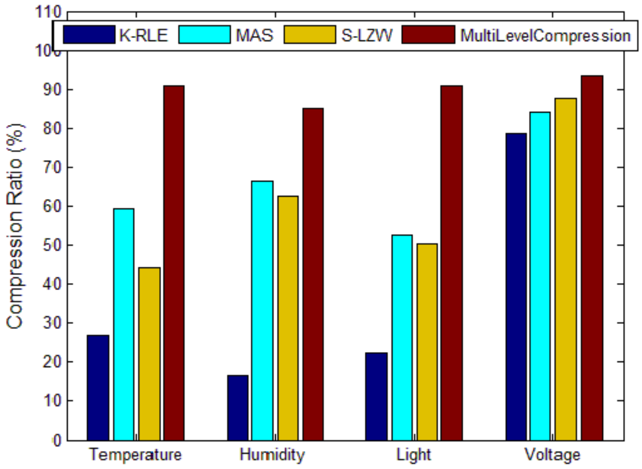


Figure 13. Compression ratios of different algorithms for different datasets

In Figure 14, we can see how various compression algorithms perform on different types of data. In this piece, we compare the suggested level compression method with three common WSN compression algorithms—K-RLE, S-LZW, and MAS.

5. CONCLUSION

In four separate ways, the work advances the cause of more efficient energy consumption during data collecting in WSN. One approach is to reduce power consumption during transmission by developing an original dual prediction framework (ASAL) that takes use of correlations in sensor data throughout time. Using a normalized-LMS filter, ASAL estimates the linear regression component of the recent data history to forecast future data. As the prediction process approaches convergence, the step size of the filter is modified accordingly. The data's dynamics dictate the adjustment of the filter's length. While maintaining little prediction error, the ASAL nLMS filter improves upon prior energy-efficient approaches.

The second approach is DEAP, which stands for data-and energy-aware clusters. To make the aggregation process more efficient, we may group nodes together based on the similarities in their data. By appointing the node with the highest energy level as leader, the cluster's energy is balanced. Using dual prediction-based reporting reduces data transmission within clusters. A direct outcome of enhanced aggregation efficiency is a decrease in data transmission between clusters. The two-tiered data reduction approach significantly improved the energy efficiency of the WSN.

The research may be enhanced by using the proposed clustering algorithms and using compressive sensing techniques. Utilizing compression-based prediction has the potential to improve the efficacy and reliability of forecasts. It is possible to improve Copest's data transmission efficiency by dissecting the system error into intra-cluster data deviation and inter-cluster data deviation. Improving the methodologies to work better with Cyber Physical Systems and the Internet of Things will make the study more useful in the real world.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

REFERENCES

[1] Ciancio, A, Pattem, S, Ortega, A &Krishnamachari, B efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm', the 5th international conference on Information processi networks, pp. 309

[2] Debono, CJ & Borg, NP reduction technique for wireless sensor networks,' IEEE International Symposium on Signal Processing and Information Technology, pp. 402-406.

[3] Janarthanan, R.; Maheshwari, R.U.; Shukla, P.K.; Shukla, P.K.; Mirjalili, S.; Kumar, M. Intelligent Detection of the PV Faults Based on Artificial Neural Network and Type 2 Fuzzy Systems. Energies 2021, 14, 6584. https://doi.org/10.3390/en14206584

[4] Baljon, M.; Sharma, S.K. Rainfall Prediction Rate in Saudi Arabia Using Improved Machine Learning Techniques. Water 2023, 15, 826. https://doi.org/10.3390/w15040826

[5] Majed Alowaidi, Sunil Kumar Sharma, Abdullah Al-Enizi, Shivam Bhardwaj," Integrating artificial intelligence in cyber security for cyber-physical systems", 'Electronic Research Archive', vol. 31, no. 4, pp-1876-1896.

[6] Garg, V., Kaur, B., Kumar, T., Alowaidi, M., Sharma, S.K. 'PIRAP: Chaotic Fuzzy Encryption (CFE) Technique and Greedy Chemical Reaction Optimization (GCRO) Algorithm-Based Secured Mobi-Cloud Framework' International Journal of Cooperative Information Systems, 2023, 32(1-2), 2250002

[7] Alzahrani, A., Alshehri, M., AlGhamdi, R., Sharma, S.K. 'Improved Wireless Medical Cyber-Physical System (IWMCPS) Based on Machine Learning' Healthcare (Switzerland), 2023, 11(3), 384

[8] Kaur, S., Kaur, N., Bhatia, K.S., ...Sharma, N.K., Sharma, S.K. 'Node localization and data aggregation scheme using cuckoo search and neural network' Expert Systems, 2023, 40(4), e13033

[9] Jitendra Singh, Maninder Singh Arora, Sunil Sharma, Jang B. Shukla. Modeling the variable transmission rate and various discharges on the spread of Malaria[J]. Electronic Research Archive, 2023, 31(1): 319- 341. doi: 10.3934/era.2023016

[10] R. AlGhamdi and S. K. Sharma, “IoT-Based Smart Water Management Systems for Residential Buildings in Saudi Arabia,” Processes, vol. 10, no. 11, p. 2462, Nov. 2022, doi: 10.3390/pr10112462. [Online]. Available: http://dx.doi.org/10.3390/pr10112462

[11] S. K Sharma* , Waseem Ahmad Khan, Cheon-Seoung Ryoo, and Ugur Duran. (2022) "Diverse Properties and Approximate Roots for a Novel Kinds of the (p,q)-Cosine and (p,q)-Sine Geometric

[12] Polynomials" Mathematics 10, no. 15: 2709. https://doi.org/10.3390/math10152709

[13] S K Sharma, M. Alwanian, M. Alowaidi, H. Alsagier (2022) “Mobile Healthcare (M-Health) based on Artificial Intelligence in Healthcare 4.0” Expert Systems DOI: https://doi.org/10.1111/exsy.13025

[14] Alshehri M* (2023) ‘Blockchain-assisted internet of things framework in smart livestock farming’ Internet of Things, Volume 22, 2023, 100739, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2023.100739.

[15] Alzahrani, A.; Alshehri, M.; AlGhamdi, R.; Sharma, S.K. (2023) ‘Improved Wireless Medical Cyber-Physical System (IWMCPS) Based on Machine Learning’ Healthcare, 11, 384. https://doi.org/10.3390/healthcare11030384

[16] Padhy S, Alowaidi M, Dash S, Alshehri M, Malla PP, Routray S, Alhumyani H. (2023) “AgriSecure: A Fog Computing-Based Security Framework for Agriculture 4.0 via Blockchain” Processes. 11(3):757. https://doi.org/10.3390/pr11030757

[17] Alshehri, M (2023) “Blockchain-assisted cyber security in medical things using artificial intelligence” Electronic Research Archive, 31(2): 708-728. doi: 10.3934/era.2023035