Maximum Frequency based Adaptive Sensing for Particulate Matter Nodes in IoT Network

by

rajashekar.reddy , Siddharth De, Sachin Chaudhari

Report No: IIIT/TR/2021/-1

Centre for Communications
International Institute of Information Technology
Hyderabad - 500 032, INDIA
June 2021

Maximum Frequency based Adaptive Sensing for Particulate Matter Nodes in IoT Network

C. Rajashekar Reddy, Siddharth De, Sachin Chaudhari
International Institute of Information Technology-Hyderabad (IIIT-H), India
Emails: rajashekar.reddy@research.iiit.ac.in,siddharth.de@students.iiit.ac.in, sachin.chaudhari@iiit.ac.in

***Abstract*—In most IoT-based monitoring applications, the data can vary at a slow rate but the variability pattern may not always be same. For example, the patterns of particulate matter (PM), one of the most dominant air pollutants, often change seasonally over a year. Therefore, having a fixed predefined sensing rate is both hard to decide and energy inefficient. This paper proposes an adaptive, non-parametric method to change the sensing rate using the maximum frequency estimate based on recent historical data. The proposed algorithm has been tested on the data collected over one year from an IoT network consisting of multiple PM sensor nodes. A performance comparison of the proposed scheme with the existing approach shows the effectiveness and performance improvement in terms of Reduction Factor (RF) and Mean Absolute Error (MAE).**

***Index Terms*—Energy Efficiency, Nyquist Rate, Particulate Matter, Periodogram, Sensing**

I. Introduction

IoT-based monitoring networks consist of distributed sensor nodes that may sense parameters such as temperature, humidity, CO_2, particulate matter (PM) concentration in the air, etc. The sensor node's limited energy capacity leads to the fact that it is one of the critical resources that need to be managed efficiently. Extension of a sensor node's lifetime is a significant challenge, especially in the case of nodes that are highly difficult to access or tedious to maintain. The lifetime of sensor nodes is also essential for the robustness and scalability of an IoT solution.

To understand the energy consumption of a sensor node, the operations of a sensor node can be divided into two major phases - standby and active [1], [2]. During the standby phase, the node is idle and can consume very low energy. The active phase includes three main energy consuming activities: sensing, computing, and transmission. Mostly, the amount of energy required for transmission is much higher than the other two activities [3]. In such cases, reducing the number of transmissions to the sink using data reduction technique significantly improves the extension of the sensor-node lifetime [3], [4]. This works very well for sensing phenomenons such as temperature and humidity, which consume significantly less energy when sensing as compared to that for transmission. So, uniformly sensing such phenomena does not consume much energy, while transmission consumes the majority of energy. In our previous work [4], it has been shown how Shewhart-based data reduction can be used to improve energy efficiency by reducing the number of transmissions. However, sensors like PM sensors have mechanical parts which use more power. For example, the Nova PM sensor SDS011 draws around 70 mA of current [5] while the DHT22 humidity sensor only draws at a maximum of 2.1 mA [6]. So, the power consumption is high for the PM sensor and energy consumed during sensing is now comparable to the energy consumed during transmission. In such a scenario, the node's energy efficiency can be improved by adaptively changing the sensing interval, which is the focus of this paper.

Generally, the data collected from monitoring applications like PM vary slowly. Some environmental phenomenon even has data patterns which change seasonally. Therefore, having a fixed and predefined sensing rate is hard to decide and highly energy inefficient. It can be easily observed that for monitoring applications like PM monitoring, the signal is usually very stable for a long time only with an addition of the sensor measurement noise. This motivates to sense the signal at far lower rates than the predefined sensing interval to improve the energy efficiency.

There have been very few works on adaptive sensing in the IoT literature [7], [8]. In [7], a sigmoid function is used to change the sensing rate. However, due to the nature of the sigmoid function, the rate of sensing rate change is equal when the parameter used lies at the extremes compared to the threshold chosen. So, the rate of change is slow.

In [8], two algorithms are implemented. The first algorithm is based on Bollinger bands [9], where the waiting time for sensing the following sample is calculated using the absolute difference between the upper and lower Bollinger bands and a dynamic estimation function. The second algorithm in [8] is based on the loss in information calculated by the vertical distance between the real-time data and the modeled data. Using the loss in information and a dynamic estimation function, the waiting time is calculated. The modeling is linear in nature. The performance comparison of the two algorithms shows that the vertical distance has better performance as compared to the one based on the Bollinger bands.

The issue with the approaches in [7], [8] is that they do not take into account the Nyquist sampling rate, which defines the lower bound on the sampling frequency [10]. According to the Nyquist sampling theorem, the sampling frequency should be more than twice the maximum frequency in the signal to avoid any reconstruction error. In this paper, we take into account the Nyquist criteria to decide the sensing frequency.

Specific contributions of this paper are

- An adaptive sensing algorithm is proposed which decides the sensing interval based on the maximum frequency estimate over a given time interval. The time interval has to be small enough to capture the seasonal variations and long enough to avoid frequent computations and change in the sensing interval.

- The proposed algorithm is tested on the data set collected from the IoT network consisting of seven PM monitoring nodes deployed inside a small educational campus in Indian city of Hyderabad. This data set corresponds to the measurements over the whole year of 2020 and consists of more than 22 million data points in total.
- The proposed approach is compared with the vertical distance based algorithm [8] in terms of Reduction Factor (RF) of sensed samples and Mean Absolute Error (MAE) for different seasons and months.

The paper is organized as follows. Section II details on IoT network deployment considered for experimental analysis followed by data cleaning and preprocessing. Section III briefly present the vertical distance based adaptive sensing algorithm. Section IV presents the proposed maximum frequency based adaptive sensing algorithm. Section V presents the comparison metric and the results while Section VI concludes the paper.

Fig. 1. Deployment and Node Locations

II. Methodology

A. Sensor Nodes

The sensor nodes are developed at IIIT-H using ESP8266 based NodeMCU microcontroller, SDS011 PM sensor, and DHT22 sensor for temperature and relative humidity as shown in [11]. The microcontroller samples the data at an interval of 15 seconds and sends it periodically via WiFi to *ThingSpeak* [12], which is a cloud-based IoT platform for storing and processing data using MATLAB. The IoT network considered in this paper is in the IIIT-H campus, Hyderabad, India. The measurement region area is 66 acres (0.267 km^2), and the data considered is for the year 2020 which has more than 22 million data points for seven nodes. Ten nodes have been deployed around the campus. Out of the ten nodes, seven nodes have been functional throughout the experimental period of one year and these seven nodes are shown in Fig. 1. The figure also shows the nodes' numbering, which will be followed for the rest of the paper.

B. Data cleaning and preprocessing

The following tasks were done to convert the seven sensor nodes' raw data into a usable dataset:

- For unbiased comparison, the data from all the months needs to be at the same sampling rate. The arrival time varied because of the additional network delay. So, the data from all nodes is resampled and grouped at intervals of 20 seconds bins.
- Due to various factors like power-cuts, network drops, and sensor failures, the data sent by sensor nodes had gaps, and some readings were zero. The longest contiguous sequence of data without zeroes was chosen for processing.
- Due to sensor errors, the data sent by the nodes contain sharp outlier peaks at some instances. The outliers need to be removed to prevent erroneous calculations. A value is considered as an outlier if it is more than three Median Absolute Deviations (MAD) away from the median of nearby values, is replaced by the median.
- For the maximum frequency based adaptive sensing approach, the power containment in the frequency domain is essential. Hence, the data has been smoothed using a moving average filter before estimating the optimal sensing frequency.

The total number of data points after cleaning and preprocessing the data is around 16 million for the seven nodes combined over the year.

III. Dynamic Sensing based on Vertical Distance

This is the algorithm proposed in [8]. A buffer $(S[1], S[2], ...S[N])$ of a fixed size N is taken. The data is initially sensed at a constant sensing rate. The first N values are pushed into the buffer. The data $x[n]$ is then modeled as a linear function between the sensed points and is given by

$$x[n] = nm + S[n], \tag{1}$$

where m is the (changing) slope of x defined by

$$m = \frac{S[N] - S[1]}{N - 1}. \tag{2}$$

Here, $S[1]$ is the oldest sensed data in the buffer and $S[N]$ is the newest. The mean absolute error Δ between $x[n]$ and

$S[n]$ for $n = 1$ to N is calculated as

$$\Delta = \frac{1}{N}\sum_{n=1}^{N} |x[n] - S[n]|, \tag{3}$$

and the estimation factor ϵ is calculated using the exponential moving average as

$$\epsilon = \frac{7}{8}\epsilon + \frac{1}{8}\Delta. \tag{4}$$

A maximum amount of waiting time t_{max} is defined which limits the value of the estimation function. The amount of waiting time is t_{wait} is defined as

$$t_{wait} = \frac{t_{max}}{1+\epsilon}. \tag{5}$$

After waiting for t_{wait} amount time, the sensor starts the active phase and senses data. After sensing the data, the node transmits and goes back to the sleep phase. The pseudocode for implementation of the vertical distance algorithm is given in Algorithm 1.

Algorithm 1 Vertical Distance Based Adaptive Sensing [8]

```
procedure VERTICALDISTANCE
    Initialize N, ε, S, and t_max
    n = 1 to N
    m = (S[N] − S[1])/(N − 1)
    Repeat = TRUE
    while Repeat == TRUE do
        x = nm + S
        Δ = (1/N) sum( abs(x − S) )
        ε = (7/8)ε + (1/8) Δ
        t_wait = t_max / (1 + ε)
        wait for t_wait
        for i = 1; i < N; i++ do
            S[i] = S[i + 1]
        S[N] = senseData()
        if wantToStop() == TRUE then
            Repeat = FALSE
```

The main problem in this method is that many float exponential and division operations are executed. So, even though the results are better than the previous algorithms, there is room for improvement by reducing the number of operations executed. Also, the algorithm requires to be implemented on the node as the sensing rate is updated with every instance as implementing on the cloud would increase the communication and power cost. The RF of the vertical distance based adaptive sensing increases with the decrease in the buffer size. The best RF for the vertical distance based dynamic sensing is obtained when the buffer size is four as the buffer size cannot reduce beyond four for this approach. This puts a constraint on the maximum amount of RF that this approach can obtain.

At the receiver, the steps for reconstruction of the data to calculate the MAE are:

- Take the sampled points along with the instances where sampling was done after the algorithm is implemented. Fill the remaining places as zeroes.
- Reconstruct the data using linear interpolation between the sampled points.

IV. Maximum Frequency Based Adaptive Sensing

The maximum frequency based adaptive sensing algorithm is a direct result of the Nyquist sampling criteria. The Nyquist criterion requires that the sampling is done at a frequency of more than twice that of the maximum frequency in the considered signal: $f_s > 2f_{max}$ [10]. However, the calculation of maximum frequency is not trivial. To determine the maximum frequency, a periodogram power spectral density is calculated and integrated using the midpoint rule [13]. $p\%$ is the percentage of total power of the signal which should be contained in the signal. The maximum frequency f_{max} is where the integrated power crosses the power containment $p\%$ threshold [14]. Based on this estimate of maximum frequency, the sensing frequency is adapted for the next time interval.

The periodogram is a non-parametric estimate of the power spectral density (PSD). For a signal $x[n]$ sampled at f_s samples per unit time, the periodogram $\hat{P}(f)$ is defined as

$$\hat{P}(f) = \frac{1}{f_s N}\left|\sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{f}{f_s} n}\right|^2, -f_s/2 < f \leq f_s/2, \tag{6}$$

where N is the number of samples considered. The summation part in (6) is Discrete Fourier Transform (DFT) of the signal and is implemented using Fast Fourier Transform (FFT) for reducing computational complexity.

Using the midpoint rule, the frequency where the integral of $\hat{P}(f)$ crosses $p\%$ power is calculated. The integral $\hat{I}(f_i)$ is defined as

$$\hat{I}(f_i) = \frac{\sum_{f=0}^{f_i} \hat{P}(f)\frac{f_s}{N}}{\sum_{f=0}^{f_s/2} \hat{P}(f)\frac{f_s}{N}} 100\%, \tag{7}$$

where $\hat{P}(f)$ is the periodogram estimate. f_{max} is defined as the value of f_i when the value of $\hat{I}(f_i)$ crosses $p\%$. A buffer of δ is added for the f_{max} to not loose the critical frequency information and avoid any possible aliasing during upsampling step in reconstruction. The f_{max} is estimated, and the sensing frequency is updated as $f_s = (2+\delta)f_{max}$ for the considered cleaned data where $\delta > 0$. δ provides a buffer over the exact threshold for f_s given by the Nyquist criterion. The pseudocode for implementation is given in Algorithm 2.

Algorithm 2 Maximum Frequency Based Adaptive Sensing

procedure MAXIMUMFREQUENCY

Obtain signal x_n of length N and f_s sensing rate

Initialize $p\%$ and δ

$\hat{P}(f) = \frac{1}{f_s N}\left|\sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{f}{f_s} n}\right|^2$

Find where $\hat{I}(f_i) = \frac{\sum_{f=0}^{f_i} \hat{P}(f)\Delta f}{\sum_{f=0}^{f_s/2} \hat{P}(f)\Delta f} 100\% > p\%$

$f_{max} = f_i$

Update $f_s = (2+\delta)f_{max}$

The value of power containment is varied in the range of 90% to 99.9%. This allows having most of the signal information retained at the lower frequency.

The maximum frequency based adaptive sensing is executed after every set amount of time using the previous historical data on the cloud. The number of data points taken needs to be sufficiently large to estimate the periodogram PSD accurately. Decreasing the length of the data considered reduces the accuracy of the estimate of the PSD. In this paper, the update has been done for the sensing frequency after every month.

The reconstruction of the newly sensed signal for calculating MAE is done using interpolation, which is the opposite of decimation [10]. The interpolation follows the following steps:

- Expand the signal to be interpolated to the initial length by inserting zeroes between the newly sensed data values which is called upsampling. The number of zeros to be inserted, called the upsampling factor d is the immediate lower integer to the ratio between initial f_s and the updated $f_s = (2+\delta)f_{max}$ resulting $d = \left\lfloor \frac{f_s}{(2+\delta)f_{max}} \right\rfloor$.
- A symmetric FIR filter that allows the original data to pass through unchanged and interpolates to minimize the mean-square error between the interpolated points and their ideal values is designed.
- The designed filter is used to interpolate the zeroes in the upsampled data by low pass filtering.

V. Results and Analysis

The comparison criteria used for analyzing the two algorithms' performance for adaptive sensing are the MAE and the RF. The RF is defined as

$$\text{RF} = \frac{N}{N_{new}}, \tag{8}$$

where N is the original number of data points and N_{new} is the number of data points after implementing the algorithm. For calculating the MAE, using the samples received after implementation, the data is reconstructed using the methods mentioned in Sections III and IV. The MAE is given by

$$\text{MAE} = \sum_{n=0}^{n=N-1} |x[n] - x_r[n]|, \tag{9}$$

where $x[n]$ is the original data and $x_r[n]$ is the reconstructed data.

For vertical distance based dynamic sensing algorithm, t_{max} is initialized as 3000 as in [8]. The buffer size N is varied from 4 to 128 by doubling N for each instance. Varying the buffer size showcases the impact of the size of the buffer on the MAE and RF. For the maximum frequency based adaptive sensing algorithm, the power containment $p\%$ has been varied from 90% to 99.9% to obtain the different reduction factors and corresponding MAE. δ is selected as 0.5 to avoid any possible aliasing during the reconstruction giving $f_s = 2.5f_{max}$. The data for the entire year of 2020 has been considered and is segregated into 12 months as well as into 3 seasons of India - Summer (March, April, May, June), Monsoon (July, August, September, October), and Winter (November, December, January, February).

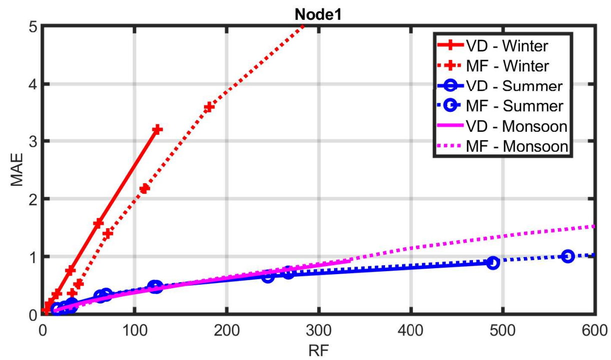
Fig. 2 shows the RF vs MAE comparison between Vertical Distance (VD) based dynamic sensing and Maximum Frequency (MF) based adaptive sensing for PM10. Similar results are followed for PM2.5 but have not been shown here because of space constraints. In Fig. 2, it can be observed that the maximum frequency based algorithm works almost always better than the vertical distance based algorithm. For Summer, it has been observed that both algorithms are at par and give similar results. However, in the Monsoon and Winter seasons, the maximum frequency based adaptive sensing algorithm gives a much higher RF with lower MAE. In Winter, PM10 is higher due to factors like humidity, temperature and festivals like Diwali due to the bursting of crackers in the residential areas in and around the campus of IIIT-H. Whereas, in Monsoon, the fluctuations in PM10 are very less due to the rains.

Another important observation from Fig. 2 is that the vertical distance algorithm stops after reaching a particular value of RF as discussed in Section III. This maximum corresponds to the buffer size of four, which is the minimum required buffer size for the vertical distance based approach. While the maximum frequency based approach does not have this limitation. We can further reduce the power containment $p\%$ till we reach the required RF or MAE. The decrease in the power containment percentage can happen until we meet the minimum points requirement for the filter used in the reconstruction.

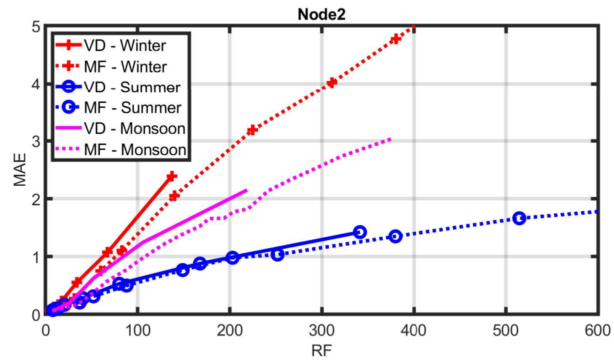
Fig. 3 presents box plots for monthly distribution of RFs for the two approaches based on vertical distance and maximum frequency for a fixed MAE $= 0.75$. It can be seen that the median of the maximum frequency algorithm is either equal or higher than the median RF of vertical distance algorithm for all the months over the year 2020 except for the month August. This shows the improvement in the RF for all the nodes on average. The environmental parameters like temperature and RH, which affect the PM concentration change over the year. The maximum frequency based adaptive sensing algorithm works better for all nodes on average, even with different months considered over the course of a year.

In Fig. 4, the average RF for MAE $= 0.75$ for the year 2020 and for the seven nodes is given as a bar chart. On average, the maximum frequency based adaptive sensing algorithm gives a better RF than the vertical distance based dynamic sensing algorithm over the year. Even though the nodes are spatially distributed around the campus of IIIT-H and have varying data patterns, the maximum frequency approach gives better performance than the vertical distance approach. The maximum difference in the performance is observed for Node7. For all the remaining nodes, the annual average reduction factors using the maximum frequency approach is higher.

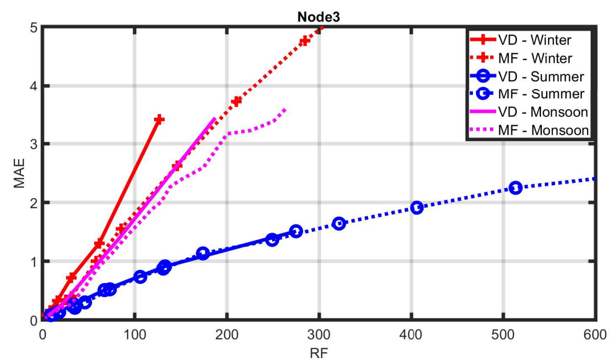
The results in Figs. 2, 3 and 4 show the improvement of performance in terms of RF and MAE using the maximum frequency based adaptive sensing approach for PM monitoring. Along with the improvement in the RF for a particular MAE,



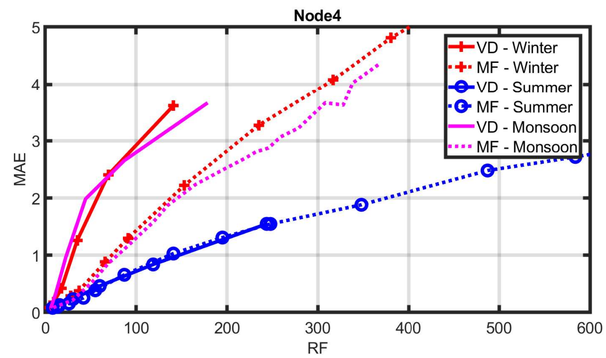
(a) Node1



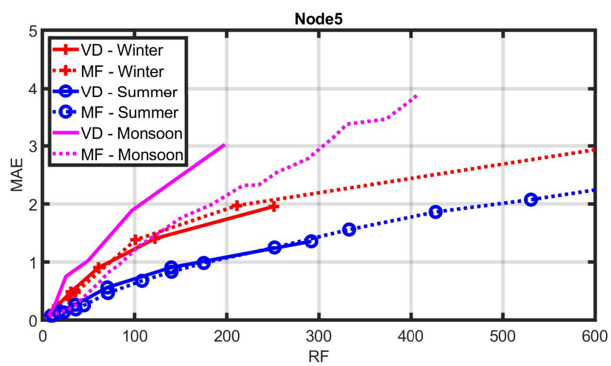
(b) Node2



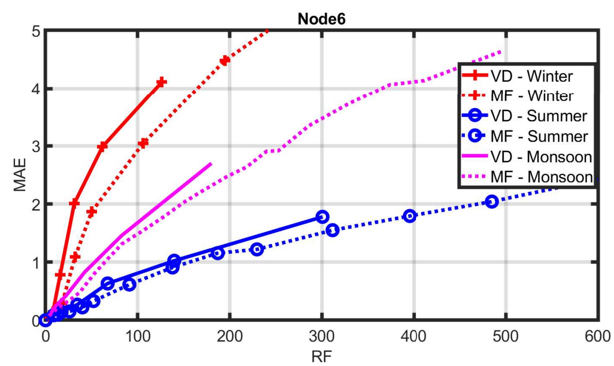
(c) Node3



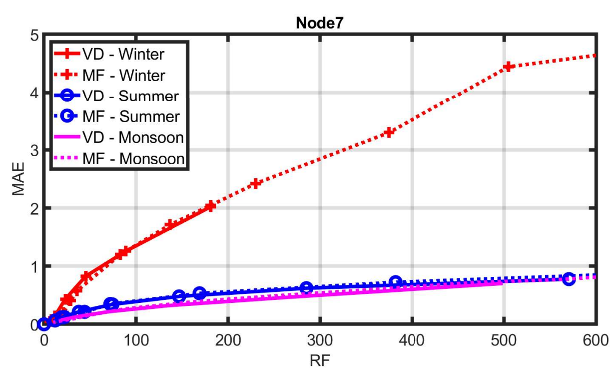
(d) Node4



(e) Node5



(f) Node6



(g) Node7

Fig. 2. RF v/s MAE - PM10 for seven Nodes

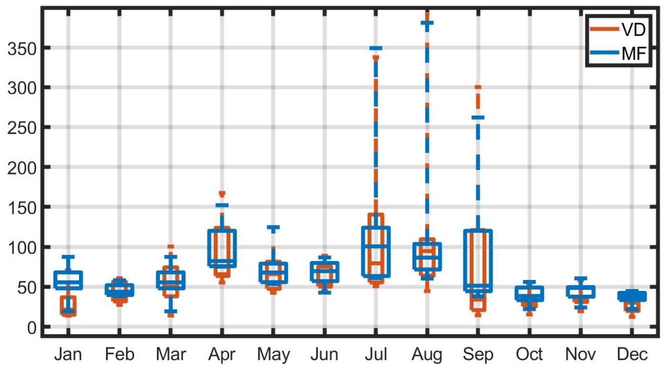


Fig. 3. PM10 sensing RF comparison over year for MAE=0.75

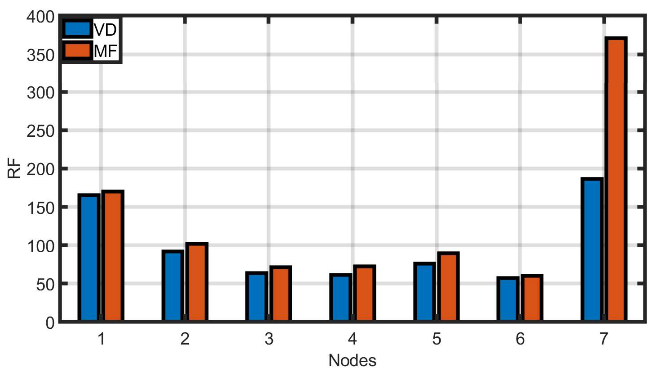


Fig. 4. PM10 sensing RF comparison for 7 Nodes for MAE=0.75

this approach enables us to offload the calculation of the new sensing interval to the cloud, and also, the number of times the calculations to be done is far less than previous approaches. In this paper, the results are shown by calculating the update to the sensing interval every month for the maximum frequency based sensing, whereas vertical distance based sensing requires updating every sample. It has also been observed that vertical distance based sensing hits a limit at a buffer size of four. In contrast, the RF and MAE trade-off can still be extended for maximum frequency based adaptive sensing by changing the power containment considered.

VI. Conclusion

In this paper, a maximum frequency based approach has been proposed for adapting the sensing interval for sensor nodes in an IoT network. The algorithm has been employed on the data collected from seven PM monitoring sensor nodes over the year. It has been tested and compared against the vertical distance based dynamic sampling algorithm. The proposed algorithm is shown to perform better than the vertical distance approach on three fronts. First, maximum frequency approach has shown better performance than vertical distance approach in terms of RF and MAE for different PM nodes across the seasons and months over the year. Second, the proposed algorithm enables us to offload the new sensing interval calculation to the cloud, and also, the number of times the estimates to be done is far less than that in vertical distance approach. Lastly, it has been shown that the RF and MAE trade-off can be extended using the maximum frequency based adaptive sensing by changing the power containment. In contrast, the vertical distance based approach hits a limit for this trade-off. Thus, the paper demonstrates that the proposed algorithm's effectiveness for designing energy-efficient IoT sensor applications by reducing the number of times the data is sensed.

Acknowledgement

This research was supported partly by National Geospatial Programme (NGP), India, under grant no. 2073 (2020), PRIF Social Incubator Program (2019) and the Ministry of Electronics and Information Technology (MEITY) under grant no. 3070665 (2020), with no conflict of interests.

References

[1] Taoufik Bouguera et al., "Energy consumption model for sensor nodes based on lora and lorawan," *Sensors*, vol. 18, pp. 2104, 06 2018.
[2] Holger Karl and Andreas Willig, *Protocols and Architectures for Wireless Sensor Networks*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005.
[3] Silvia Santini and Kay Uwe Römer, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proceedings of INSS*. 2006, Transducer Research Foundation.
[4] A. Shastri, V. Jain, S. Chaudhari, S. S. Chouhan, and S. Werner, "Improving accuracy of the shewhart-based data-reduction in iot nodes using piggybacking," in *IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019.
[5] *SDS011 Nova Sensor Specifications*, accessed 18 Mar. 2021, http://www.inovafitness.com/en/a/chanpinzhongxin/95.html.
[6] *DHT22 Sensor Specifications*, accessed 18 Mar. 2021, http://www.adafruit.com/datasheets/DHT22.pdf.
[7] Tongxin Shu et al., "An energy efficient adaptive sampling algorithm in a sensor network for automated water quality monitoring," *Sensors*, vol. 17, no. 11, 2017.
[8] U. Kulau et al., "Dynamic sample rate adaptation for long-term iot sensing applications," in *IEEE 3rd World Forum on Internet of Things (WF-IoT)*, 2016, pp. 271–276.
[9] Oliver Douglas Williams, "Empirical Optimisation of Bollinger Bands for Profitability," 2006.
[10] Li Tan and Jean Jiang, *Digital Signal Processing: Fundamentals and Applications*, Academic Press, Inc., USA, 2nd edition, 2013.
[11] C. Rajashekar Reddy et al., "Improving Spatio-Temporal Understanding of Particulate Matter using Low-Cost IoT Sensors," *IEEE PIMRC*, 2020.
[12] *ThingSpeak*, accessed 18 Mar. 2021, https://thingspeak.com/.
[13] Edwin "Jed" Herman Gilbert Strang, *Calculus Volume 2*, OpenStax, Houston, Texas, 2016.
[14] Sean A. Fulop and Kelly Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 360–371, 2006.