Air Quality Monitoring and Analysis Network

Somil Singhai, Akshat Baranwal, Ankush C. Jahagirdar, Karunesh K. Gupta*
BITS Pilani University - Pilani, India
*kgupta@pilani.bits-pilani.ac.in

***Abstract*—With the advent of technology, more and more aspects of our lives are being digitized. Hence, eventually, there will be a need for sensor-based networks for creating sustainable living conditions. This paper aims to describe the implementation and application of an Air Quality monitoring and analysis network(AQMAN) capable of monitoring different air quality parameters which could then be used to predict the sustainability of a locality at the expense of precision. The network employs various Machine Learning algorithms for forecasting the parameters on multiple time granularity. A method for constructing a Geospatial graph of the parameter's concentration has also been discussed in the later sections. The paper takes a more pragmatic approach of making the system cheaper, reliable, scalable and accessible.**

***Index Terms*—Soft Sensors, Air Quality Index(AQI), Cloud, IoT, LSTM**

I. Introduction

It is estimated that people, on average spend more than 90% of their time indoor. This comprises of the time spent at work, school and of course, at home. Because of the amount of time we spend inside [1], the quality of the air we breathe should be a major concern for all of us. The Environmental Protection Agency(EPA) has ranked indoor air pollution as the most concerning environmental dangers, we all face daily [1].Hundreds of harmful chemicals are released by household cleaning agents, personal care products, paint and various solvents regularly [2].

The main atmospheric pollutants include organic compounds, carbon compounds, toxic gases(nitrogen dioxide, sulphur dioxide, ozone, etc) and particulate matter(PM) produced by human activities. At present, various agencies, around the world have installed and are operating large fixed Air Quality Monitoring (AQM) stations. These agencies provide accurate and reliable information but with a meagre temporal and spatial resolution, which is unfit for data analysis at personalized scale.

The paper tries to establish a development process of a robust data acquisition system in conjunction with applying various Machine learning techniques, specifically time-series prediction and fuzzy-logic for decision making. The design of the system is completely modular, which makes adding or removing the nodes, very smooth. The primary contributions of this work are as follows:

1) Design of an affordable modular hardware which supports different compatible sensors that can be added depending on the use case and working environment.
2) Each node of our network supports Over-the-air (OTA) updates. To the author's best knowledge, this work is the first in the concerned field to support such a feature.

The remaining paper is organized as follows: Section II discusses related works about the air quality monitoring solutions. Section III presents the overall network architecture while Section IV concludes the paper, suggesting future works and author's expectations.

II. Related Works

Various Researchers have published their work in this field. From the hardware perspective, many authors have used ZigBee technology for constructing the nodes such as in [3]. A sensor network working on star mesh configuration has been proposed in [4]. Most of the solutions such as in [5]–[7] employ a gateway node for collecting and sending the data. This induces an extra source of complications and also increases the cost. [8] have proposed a Modular Sensor System (MSS) architecture and Universal Sensor Interface (USI), with a modular design in a sensor node.

For the computing and storage part, we have used cloud computing. This technology virtually provides us with infinite computing power and storage. In fact, many cloud services such as [9], [10] and [11] have started offering integrated IoT solutions along with cloud services. In this paper, we have used one such service along with our customization to comply with our needs.

A number of studies in the literature have proposed solutions for predicting AQI. They employ a combination of Neural Networks [12], [13] and Empirical approaches which can be broadly categorized into deterministic and statistical methods. [14] proposes a hybrid between linear regression and Artificial Neural Network for forecasting. [15] also proposes a hybrid combination of Ensemble Empirical mode decomposition(EEMD) and Mirror method(MM) which provides robust results.

III. System Architecture

The system is divided into three layers - the sensor network, Cloud services(specifically Google Cloud services - GCP) and end user services as shown in Fig. 1. The following subsections discuss this in detail.

A. Sensor Layer

1) Networking and Communication: We assessed various communication components for our purpose such as Raspberry Pi 3, Arduino UNO along with ESP8266 WiFi node, ZigBee

Funded by Academic-Undergraduate Studies Division, BITS-Pilani

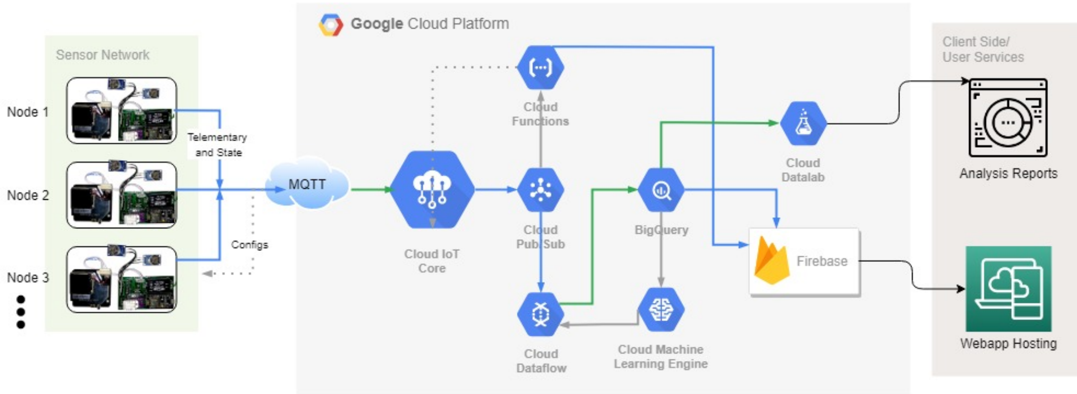

Fig. 1. System Architecture consisting of three layers - The Sensor Network, Google Cloud services and Client Side Services

Xbee, and finally NodeMCU. Table.I shows various criteria of comparisons, based on which the selection is done.

Transmission Reliability (TR) is calculated by finding the number of successful data update over 1000 update calls. We found UNO+ESP8266 to be highly unreliable. Ultimately we decided to go with NodeMCU due to its value for money.

2) Micro Controller: NodeMCU served both as a controller(MCU) as well as a communication segment. It is equipped with a Flash memory of 4MB and SRAM of 64KB, at the same time supporting IEEE 802.11 b/g/n WiFi protocol. It has a single Analog pin which posed a significant problem. We solved this problem by using a 8-pin multiplexer. Three Digital pins from NodeMCU were used as selection pins for the multiplexer. The Controller is powered via. external power source. This takes a toll on the portability of the node but solves the power consumption problem posed by the sensors.

Each node supports OTA updates which is critical since it allows us to remotely fix any problem. It also helps in software calibration and resolving security issues.

3) Sensors: All environmental sensors are connected to the MCU via different protocols such as Universal Asynchronous Receiver/Transmitter(UART), 1-wire bus, and Serial Peripheral Interface(SPI). Table.II shows the various sensors and their ranges, used in each of our node.
Gas Sensors: Each node supports a number of MQ gas sensor series. But we focused on the major pollutants found indoor and thus chose MQ2 and MQ135 for our purpose. These sensors have a good precision and are cheap. Unfortunately, the inherent disadvantage lies in their design. Being oxide based, they require about 24-48 hours of preheating time, before they start providing sufficiently precise measurements. The second disadvantage is their power consumption. Each MQ sensor consumes about 120 mA of current, which results in a overall high power consumption and also make powering the node using a battery, impractical.

4) Mechanical Structure and PCB: The casing of node is 3D printed using PLA plastic. The experimentation is done using sample sensors at various levels upto the height of an average person to mimic real conditions. Fig. 2 shows the overview of core functional circuit board which is designed in Easy EDA Designer.

B. Cloud Services

The data flow of main routines has been shown in the Fig. 1. The communication in the system is done using the Message Queuing Telemetry Transport(MQTT) protocol. A number of works have already used MQTT for communication between the server and Sensor network such as in [13], [16]. After acquiring the data from all the sensors, the MCU publishes the measurements to each of the subjects. These subjects are later subscribed by various services such as Front End and Cloud Functions. In the following sub sections, the role of each service is briefly explained.

1) Cloud IoT core: The MCU collects the sensor data and combines them into the message packet every t second (where t is the time interval between every measurement). GCP's Cloud IoT core service is responsible for registering the devices and

TABLE I
COMPARISON OF VARIOUS MCUS

Component	Range	Transmission Reliability	Cost (Rupees)
Raspberry Pi3	high	98%	2900
Arduino UNO + ESP8266	low	60%	680
Xbee Zigbee	medium	high	980
NodeMCU with ESP8266	low	93%	325

TABLE II
LIST OF SENSORS USED IN EACH NODE

Component	Quantity measured	Measurement Range
DHT22	Temperature & Humidity	-40°C∼+125°C RH:0∼100 ±2.5%
Nova PM Sensor SDS011	PM2.5 & PM10	0.0-999.9 μg/m^3
MQ2	Methane, Butane, LPG, smoke	200∼10000ppm
MQ135	Benzene, Alcohol, smoke	10∼300ppm NH3 10∼1000ppm C6H6

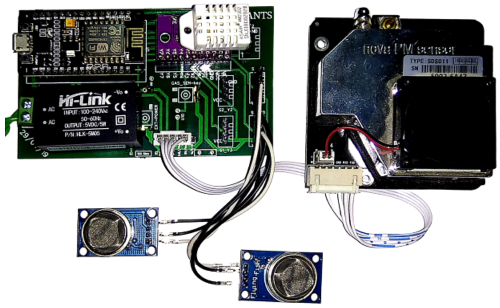

Fig. 2. A Sensor Network Node without the casing

Fig. 3. Geospatial Distribution of Concentration of a Parameter

also segregating the values propagated via. different nodes. In case the node is unable to transmit the values, say in case of a network outage, the node stores some values in its flash memory with an increased time interval.

GCP Cloud Pub-sub is a publisher and subscriber service which acts as a MQTT broker and allows us to ingest data in real-time. It allows communication between cloud function and cloud data flow and allows us to manage the various topics and their subscribers. Initially, when PHP was chosen it lead to high consumption of server resources such as memory, giving rise to high latency. This lead to compromised accuracy, reliability, and promptness of the data. Finally NodeJS is used for programming the Back End. It not only provided a more friendly programming paradigm but it could also handle much more concurrent user requests than PHP.

2) Cloud Functions: Cloud functions is used to manage various routines which are initiated whenever there are certain events (such as switching on of a new node, abnormal values in the database, etc.). It also responds to https calls from Firebase for certain purposes such as switching off of certain nodes and registering a new node. Basically it is responsible for almost all of the asynchronous tasks happening in the system.

3) Cloud Dataflow: GCP's Cloud Dataflow is a batch and real-time data processing service [17]. It is capable of accepting data stream from cloud pub-sub and storing them to BigQuery. It also supports SQL queries which may be used by Firebase in some cases to serve on-demand historical data about various air quality parameters.

4) Data Analysis: Cloud Machine Learning Engine supports dynamic training and also allows us to manage our model versions. The model is trained using the reference nodes(in controlled environment). The model constructed is then used to provide forecasting for the user nodes. Later we intend to use other nodes as well for training purposes. A very good work has been done in [13] on utilising LSTM for air quality prediction. We applied the same method for forecasting. The paper shows that stochastic factors are more prevalent in determining the concentration of pollutant and hence doing a multivariate analysis along with temperature, humidity does not significantly benefit the model in terms of accuracy. This module also serves as a source of warnings and information pertaining to the air quality by using fuzzy logic. So for example, a warning is issued if the concentration of multiple parameters reaches above a threshold value.

C. End User Services and Results

Firebase serves as a platform for Mobile and web application development. Details of Front end won't be described in much detail since its development has become quite mainstream and has already been covered by a lot of works in the past. The web app supports visualization of historical on various time-scales such as hourly, daily and weekly. A novel ability of the App is to show the concentration of pollutants in a '3D-Bar Map', with higher bars indicating a higher concentration. However, this service still needs to be integrated with the main system. For now, we have used an artificially generated data to make a plot as shown in Fig. 3. The Location is Mumbai, India.

IV. Conclusion

The proposed AQMAN system is not intended to replace the large scale monitoring stations, rather it is proposed to use this network as a complementary source of information. Aimed at study on the problem of urban air pollution, the work develops a real-time sensor system capable of monitoring the air quality and providing health related insights based on the data. We have also tried to show the vast potential of cloud technology in building these reliable and highly scalable solutions. Considering the fact that real applications of this system require multiple concurrent devices online, the performance in the monitoring system is evaluated in terms of overall performance, latency, and accuracy of data. Also, by carefully selecting the hardware components as well as software configuration, the solution is highly effective and has a wide potential application.

Acknowledgment

This Research project is supported by Academic-Undergraduate Studies Division of Birla Institute of Technology and Science University, Pilani.

References

[1] EPA. (2017) Indoor air quality. [Online]. Available: https://www.epa.gov/report-environment/indoor-air-quality

[2] I. of Medicine (États-Unis). Committee on the Effect of Climate Change on Indoor Air Quality and P. Health, *Climate change, the indoor environment, and health*. National Academies Press, 2011.

[3] W. Fuertes, D. Carrera, C. Villacs, T. Toulkeridis, F. Galrraga, E. Torres, and H. Aules, "Distributed system as internet of things for a new low-cost, air pollution wireless monitoring on real time," in *2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, Oct 2015, pp. 58–67.

[4] J. Lozano, J. I. Suárez, P. Arroyo, J. M. Ordiales, and F. Álvarez, "Wireless sensor network for indoor air quality monitoring," 2012.

[5] P. Arroyo, J. L. Herrero, J. I. Surez, and J. Lozano, "Wireless sensor network combined with cloud computing for air quality monitoring," *Sensors*, vol. 19, no. 3, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/3/691

[6] B. Guanochanga, R. Cachipuendo, W. Fuertes *et al.*, "Real-time air pollution monitoring systems using wireless sensor networks connected in a cloud-computing, wrapped up web services," in *Proceedings of the Future Technologies Conference (FTC) 2018*. Cham: Springer International Publishing, 2019, pp. 171–184.

[7] J. Esquiagola, M. Manini, A. Sequeira Aikawa, L. Yoshioka, and M. Zuffo, "Monitoring indoor air quality by using iot technology," 08 2018.

[8] W.-Y. Yi *et al.*, "A modular plug-and-play sensor system for urban air pollution monitoring: Design, implementation and evaluation," *Sensors (Basel, Switzerland)*, vol. 18, no. 1, p. 7, Dec 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29271952

[9] Microsoft. Microsoft azure services. [Online]. Available: https://azure.microsoft.com/en-in/

[10] Amazon. Amazon web services. [Online]. Available: https://aws.amazon.com/

[11] Google. Google cloud platform. [Online]. Available: https://cloud.google.com/

[12] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, "Deep learning architecture for air quality predictions," *Environmental Science and Pollution Research*, vol. 23, 10 2016.

[13] . Kk, M. U. imek, and S. zdemir, "A deep learning model for air quality prediction in smart cities," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec 2017, pp. 1983–1990.

[14] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data." KDD 2015, August 2015. [Online]. Available: https://www.microsoft.com/en-us/research/publication/forecasting-fine-grained-air-quality-based-on-big-data/

[15] J. Zhu *et al.*, "A hybrid forecasting approach to air quality time series based on endpoint condition and combined forecasting model," *International journal of environmental research and public health*, vol. 15, no. 9, p. 1941, Sep 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/30200597

[16] R. K. Kodali and B. S. Sarjerao, "Mqtt based air quality monitoring," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dec 2017, pp. 742–745.

[17] Google. Dataflow. [Online]. Available: https://cloud.google.com/dataflow/