2020 International Conference on Computational Science and Computational Intelligence (CSCI)

A Low Cost LoRa-based IoT Big Data Capture and Analysis System for Indoor Air Quality Monitoring

Matthew Meli (Contact Author)
Department of Microelectronics and Nanoelectronics
University of Malta
Msida, Malta
matthew.meli@um.edu.mt

Edward Gatt
Department of Microelectronics and Nanoelectronics
University of Malta
Msida, Malta
edward.gatt@um.edu.mt

Owen Casha
Department of Microelectronics and Nanoelectronics
University of Malta
Msida, Malta
owen.casha@um.edu.mt

Ivan Grech
Department of Microelectronics and Nanoelectronics
University of Malta
Msida, Malta
ivan.grech@um.edu.mt

Joseph Micallef
Department of Microelectronics and Nanoelectronics
University of Malta
Msida, Malta
joseph.micallef@um.edu.mt

***Abstract*— This paper presents a low cost LoRa-based IoT big data capture and analysis system for indoor air quality monitoring. This system is presented as an alternative solution to expensive and bulky indoor air quality monitors. It enables multiple low cost nodes to be distributed within a building such that extensive location-based indoor air quality data is generated. This data is captured by a gateway and forwarded to a cloud-based LoRaWAN network which in turn publishes the received data via MQTT. A cloud-based data forwarding server is used to capture, format and store this big data on a cloud-based document-oriented database. Cloud-based services are used for data visualization and analysis. Periodic indoor air quality graphs along with air quality index and thermal comfort index heat maps are generated.**

Keywords—IoT, LPWAN, Indoor Air Quality, Big Data, LoRa

I. INTRODUCTION

The air one breathes is becoming increasingly polluted mostly due to urbanization and industrialization. One often has the perception that air pollution is present only outdoors and rarely thinks about indoor air pollution despite spending 90% of the time indoors. Some indoor pollutants originate from the outside, but most of them are released inside the building. Common sources are furniture, cooking, heating, or cleaning. Air pollutants can build up quickly indoors especially if the ventilation rate of the building is inadequate. This can lead to several health conditions such as the Sick Building Syndrome and the Building Related Illness. Indoor air quality (IAQ) monitoring is hence essential to provide information which enables the adequate control of the building ventilation through building management systems (BMS). Such improved control can help to improve the energy performance of the building but also ensures that the indoor air is adequately filtered and diluted with outdoor air to provide acceptable living conditions [1].

IAQ monitors are often bulky and expensive devices. This limits their usability in most buildings. Furthermore, buildings equipped with IAQ monitors typically only have a few IAQ devices. This results in limited location-based IAQ information [2]. As a solution, this work presents the design and implementation of a low cost LoRa-based Internet of Things (IoT) big data capture and analysis system for IAQ monitoring. The low cost IoT system enables multiple nodes to be distributed within the building while generating extensive location-based IAQ data. The design of the IAQ monitoring system is presented along with the testing results of an implemented proof of concept system.

II. BACKGROUND

The main building block of an IoT system is the wireless communication protocol. This is normally selected based on the application requirements. In this work LoRa, which is a low power wide area network (LPWAN), was identified as being most suitable due to the large link budgets available and its low power consumption. Despite having low data rates, LPWAN technologies outclass IEEE 802.11 and IEEE 802.15 protocols in terms of range and power consumption. Nonetheless, the maximum data rate achieved by LoRa, 50 Kbps, is more than enough for such an IAQ application. Furthermore, LoRa through the use of configurable bandwidths and spreading factors can be optimised by finding an optimal balance between data rate, range and power [3, 4].

A. Network Architecture

The LoRa network architecture consists of a star of stars topology with IoT nodes being able to communicate with one or several gateways. A router forwards the data received by the gateway to a LoRaWAN server. The LoRaWAN server is responsible for device management and provides application programming interfaces (API) which enable data access to the data analysis tools [4].

TABLE I. LOW COST LORA TRANSCIEVER MODULES

Transceiver Module	**Tx Power (dBm)**	**Rx Sensitivity (dBm)**	**Current**			**Cost**
			Tx (mA)	***Rx (mA)***	***Sleep (µA)***	
CMWX1ZZABZ	14	-135.5	47	23.6	1.4	€15
iM88A-XL	10/13/15	-138	25/29/38	11.2	1.4	€12
RAK811	14	-130	30	5.5	7.2	€10
RN2483	14	-148	28/39	14.2	2	€8
RFM95W	7/13/17/20	-148	20/29/87/120	10.3	0.2	€6

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00070

Low cost LoRa transceiver modules are presented in Table I. These transceivers are based on the Semtech SX1276 transceiver and have similar specifications. Similarly, low cost LoRa gateways are presented in Table II. Most of them are based on the Semtech SX1301 baseband processor and a Semtech SX1257 transceiver providing 8 channels with concurrent reception capabilities.

TABLE II. LOW COST LORA GATEWAYS

Gateway	Ethernet/Wi-Fi/ LTE/GPS	Rx Channels	Rx Sensitivity (dBm)	Cost
TTN Gateway	Yes/Yes/No/Yes	8	N/A	€280
WLRGFM-100	Yes/Yes/No/No	8	-142	€180
RAK831 + Raspberry Pi	Yes/Yes/No/Yes	8	-142	€175
RAK7258	Yes/Yes/No/No	8	-142	€140
LG02	Yes/Yes/No/No	2	-148	€80
TTN Indoor	No/Yes/No/No	8	-140	€65
LG01-S	Yes/Yes/No/No	1	-148	€50

Several LoRaWAN network servers are available on the market with companies like Loriot and Senet providing their proprietary solutions. In contrast, free open source LoRaWAN network servers are limited since they have only two main options: The Things Network and Chirp Stack. The specifications of these two LoRaWAN network servers are presented in Table III. Both of them allow the deployment of an unlimited number of gateways and IoT nodes. The main difference is that The Things Network is public while Chirp Stack is private.

TABLE III. FREE OPEN SOURCE LORAWAN NETWORK SERVERS

LoRaWAN Server	Open Source	Network Type	Maximum Number Gateways/Devices	API Support
The Things Network	Yes	Public	Unlimited/Unlimited	Yes
Chirp Stack	Yes	Private	Unlimited/Unlimited	Yes

B. Data Storage

Big data is stored in databases which are typically either of the relational type or the document-oriented type. Relational databases are organized into tables and require data normalization to ensure optimal results. Normalization requires the use of additional tables, joins, keys, and indexes resulting in the database being hard to scale in web and cloud big data applications. In contrast, document-oriented databases do not make use of SQL and tables but instead rely on structured JSON like documents. This results in a distributed and horizontally scalable database enabling faster and easier data access by APIs [5].

The most utilized document-oriented database is MongoDB. MongoDB is specifically built for cloud applications and it allows API integrations for data analysis. The MongoDB Atlas service allows databases to be deployed on shared clusters on the cloud. The service is for free up to 512MB of data but can be easily scaled up with dedicated cloud clusters. MongoDB Atlas provides a Web UI through which real-time performance metrics and data management and security tools can be accessed [5, 6].

C. IoT Data Messaging Protocols

Cloud-based IoT services make use of lightweight messaging protocols in order to share data between them. The most established IoT data messaging protocols are: Message Queuing Telemetry Transport Protocol (MQTT), Constrained Application Protocol (CoAP), Advanced Message Queuing Protocol (AMQP) and Hyper Text Transport Protocol (HTTP). The specifications of these protocols are summarized in Table IV. MQTT provides the smallest header size and has the most extensive set of QoS options. [6, 7].

TABLE IV. IOT MESSAGING PROTOCOL SPECIFICATIONS [7]

Criteria	MQTT	CoAP	AMQP	HTTP
Architecture	Client/Broker	Client/Server Client/Broker	Client/Server Client/Broker	Client/Server
Header Size	2 Byte	4 Byte	8 Byte	Undefined
QoS Reliability	At most once At least once Exactly once	At most once At least once	At most once At least once	Limited
Transport Protocol	TCP	UDP, SCTP	TCP, SCTP	TCP
Security	TLS/SSL	DTLS, IPSec	TLS/SSL, IPSec, SASL	TLS/SSL
Default Port	1883/8883	5683/5684	5671/5672	80/443
Licensing	Open source	Open source	Open source	Free

III. DESIGN AND IMPLEMENTATION

A block diagram of the designed low cost IoT big data capture and analysis system shown in Fig 1. The IoT system is made up of the LoRa network architecture which gathers IAQ data from the distributed sensor nodes and the cloud-based services which perform data forwarding, storage, visualization, and analysis.

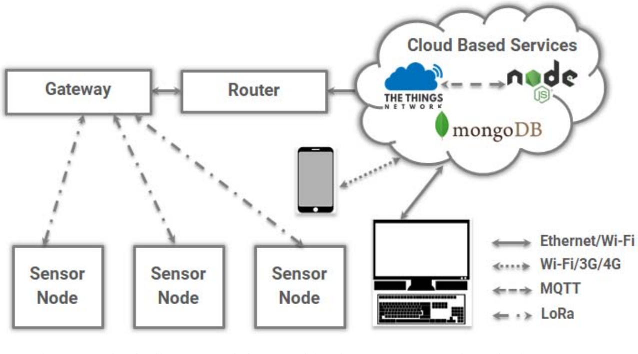


Fig. 1. Block diagram of the IoT big data capture and analysis system

A. LoRa Network Architecture

A low cost prototype IAQ IoT sensor node was specifically designed for this application. The specifications of this IoT node are presented in Table V. The node has extensive sensing capabilities and is equipped with an RFM95W LoRa transceiver. This transceiver features a low power consumption and is one of the cheapest found on the market.

TABLE V. IAQ IOT SENSOR NODE SPECIFICATIONS

Sensing Ability	Additional Interfaces	Wireless Interface	Battery Life	Cost
Temperature, Humidity, Pressure, Carbon Dioxide, Volatile Organic Compounds, Particulate Matter	UART, I2C, SPI, Analogue	LoRa	14 months	€175

Following the selection of the LoRa transceiver, a low cost LoRa gateway was selected. The selected LoRa gateway is the RAK7258 gateway. This is one of the most affordable LoRa gateways which has both Ethernet and Wi-Fi connectivity. This gateway provides OpenWRT software which allows Web UI monitoring of LoRa parameters such as the duty cycle, traffic, received signal strength intensity (RSSI) and signal to noise ratio (SNR).

The gateway is hosted on the The Things Network LoRaWAN server. The Things Network is also responsible for the device management. Devices on this server are added using either over the air activation (OTAA) or activation by personalization (ABP). OTAA is more secure than the ABP as the device address is dynamically assigned and security keys are negotiated during the join procedure.

B. *Data Storage and Analysis*

The IoT data being generated is in JSON like format and therefore MongoDB Atlas, a cloud-based document-oriented database, is used. The database was deployed on a shared cloud cluster in a three-replica set. This results in the system having more capacity for distributed read operations while being more redundant against system failures. The database is split in three collections: metadata, air quality and location. Table VI shows the document fields present in each collection.

TABLE VI. DATABASE COLLECTION DOCUMENT FIELDS STRUCTURE

Metadata Collection	Air Quality Collection	Location Collection
_id (Index)	_id (Index)	_id (Index)
Application ID	Device ID	Device ID
Device ID	Floor	Air Quality Index
Hardware Serial Number	Block	Thermal Comfort Index
Port	Room	Main Pollutant
Frame Counter	Temperature	X Position
Raw Payload	Relative Humidity	Y Position
Battery Voltage	Pressure	Block
Frequency	Volatile Organic Compunds	Level
Modulation	Carbon Dioxide	Date and Time
Data Rate	Particulate Matter PM1.0	
Coding Rate	Particulate Matter PM2.5	
RSSI	Particulate Matter PM4.0	
SNR	Particulate Matter PM10	
Air Time	Typical Particle Size	
Channel	Air Quality Index	
Gateway ID	Thermal Comfort Index	
Date and Time	Main Pollutant	
	Date and Time	

MongoDB Charts, another cloud-based service is utilized for data analysis. Charts showing the periodic mean or raw levels of IAQ data are generated using this service. Filters can be applied on these charts to show specific data ranges and isolate specific sensor nodes. Air quality index (AQI) and thermal comfort index (TCI) heat maps were generated by mapping the building and the locations of the IAQ IoT nodes using a coordinate system.

C. *IoT Data Forwarding Server*

The data from the IoT nodes is formatted using the Cayenne low power payload (LPP) protocol before being sent over LoRa. This protocol conforms to the Internet Protocol for Smart Objects (IPSO) Guidelines. Once uploaded on the Things Network, the IAQ data is published over an MQTT bridge. A node.js data forwarding server is deployed to capture the MQTT messages and store them in one of the three MongoDB Atlas collections shown in Table VI.

Before storing the IAQ data, the node.js data forwarding server formats the data and generates an AQI and TCI. The main pollutant contributing to the generated AQI is also recorded. The implemented AQI and TCI are based on [8]. Tables VII and VIII describe the AQI and TCI respectively. Whenever an IAQ IoT sensor node transmits IAQ data, the AQI and TCI values in the location collection are updated accordingly to generate the AQI and TCI heat maps.

TABLE VII. AIR QUALITY INDEX FOR INDOOR AIR QUALITY MONITORING

Pollutant	Good (1)	Moderate (2)	Unhealthy (3)	Hazardous (4)
PM10 (μg/m^3)	0-20	21-150	151-180	181-600
PM2.5 (μg/m^3)	0-15	16-40	41-65	66-500
CO2 (ppm)	340-600	601-1000	1001-1500	1501-5000
TVOC (ppb)	0-87	88-261	262-430	431-3000

TABLE VIII. THERMAL COMFORT INDEX FOR INDOOR AIR QUALITY MONITORING

Pollutant	Good (1)	Moderate (2)	Unhealthy (3)	Hazardous (4)
Temperature (^{0}C)	20-26	17-19.9, 26.1-29	7-16.9, 29.1-39	0-6.9, 39.1-45
Relative Humidty (%)	40-70	70.1-80	80.1-90	90.1-100

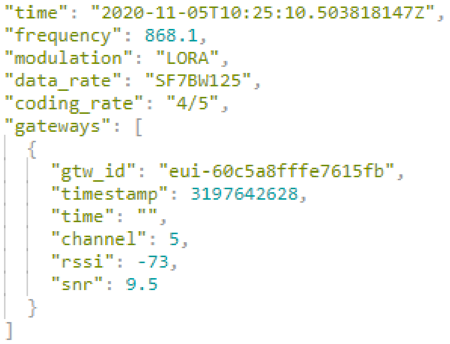
IV. RESULTS

In order to test the low cost IoT big data capture and analysis system developed, ten prototype IAQ IoT nodes were deployed alongside a LoRa gateway within an enclosed office building. The selected building testbed, shown in Fig. 2, forms part of the University of Malta campus. The building has four floors and is made up of a concrete sub-structure with a steel upper structure. The upper structure is fully enclosed by glass and the building has no windows. The enclosed building has a BMS system part of which is responsible for controlling the ventilation and air conditioning system. This makes this building an ideal testbed for the developed IAQ big data capture and analysis system.

Fig. 2. Building testbed for the IoT big data capture and analysis system

The propagation properties of LoRa within the building and the correct operation of the LoRa network architecture were first tested. This was done to ensure that an optimum coverage throughout this challenging building structure is achieved whilst guaranteeing that that messages from the registered devices were being successfully received on the LoRaWAN network.

A single gateway with a sensitivity of -142 dBm was deployed in a central location within the building and IoT sensor nodes were moved along the different locations within the building. Several LoRa packets were sent from each individual location with the corresponding RSSI and SNR being measured each time using the metadata generated by messages sent to The Things Network as shown in Fig. 3.



```
"time": "2020-11-05T10:25:10.503818147Z",
"frequency": 868.1,
"modulation": "LORA",
"data_rate": "SF7BW125",
"coding_rate": "4/5",
"gateways": [
  {
    "gtw_id": "eui-60c5a8fffe7615fb",
    "timestamp": 3197642628,
    "time": "",
    "channel": 5,
    "rssi": -73,
    "snr": 9.5
  }
]
```

Fig. 3. Metadata of a LoRa packet received on The Thing Netowrk

During this test, the LoRa parameters of the transceiver were set according to Table IX. The spreading factor was set to the lowest possible setting (7). This was done to ensure that the worst-case coverage and range scenarios are tested.

TABLE IX. LOW RANGE, HIGH DATA RATE LORA PARAMETERS

Spreading Factor	Bandwidth	Coding Rate	Tx Power	Minimum SNR	Minimum RSSI
7	125 KHz	4/5	14 dBm	-7.5 dB	-142.5 dBm

Fig. 4 and 5 show the test RSSI and SNR results achieved using this LoRa configuration over the different locations within the building. Optimum coverage is achieved across most of the building using the single gateway. The only packets lost due to the minimum SNR limit being exceeded were at the edges of the basement, level -2. These locations are the furthest from the gateway and hence a second gateway was required for complete coverage of the building.

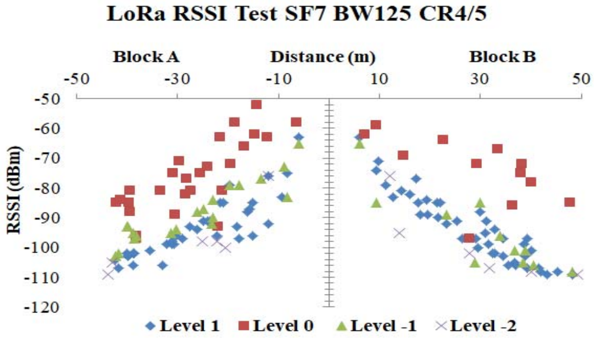


Fig. 4. Indoor LoRa RSSI test

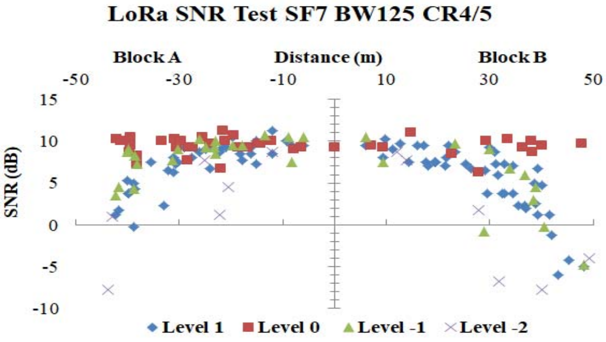
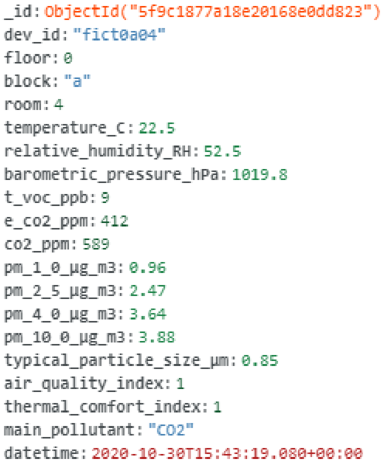


Fig. 5. Indoor LoRa SNR test

Once optimum coverage throughout the whole building was ensured and the correct operation of the LoRaWAN network was verified, the functionality of the MQTT data forwarding server was confirmed. This was done by ensuring the successful storage of the formatted JSON data on the cloud-based database. Fig. 6 shows a typical formatted IAQ record stored in the Air Quality database collection.



```
_id: ObjectId("5f9c1877a18e20168e0dd823")
dev_id: "fict0a04"
floor: 0
block: "a"
room: 4
temperature_C: 22.5
relative_humidity_RH: 52.5
barometric_pressure_hPa: 1019.8
t_voc_ppb: 9
e_co2_ppm: 412
co2_ppm: 589
pm_1_0_µg_m3: 0.96
pm_2_5_µg_m3: 2.47
pm_4_0_µg_m3: 3.64
pm_10_0_µg_m3: 3.88
typical_particle_size_µm: 0.85
air_quality_index: 1
thermal_comfort_index: 1
main_pollutant: "CO2"
datetime: 2020-10-30T15:43:19.080+00:00
```

Fig. 6. IAQ record in the Air Quality database collection

Following the functional verification of all the IoT systems IAQ data was collected and data analysis on the database was then performed using mongoDB charts. Fig. 7 shows daily particulate matter data being collected by one of the nodes while Fig. 8 shows monthly volatile organic compound data. In addition, yearly or specific time frame graphs can also be generated with the use of filters. Moreover, Fig. 9 shows that apart from visualizing IAQ data from a single IoT node, the system can be configured to show data from multiple nodes. Once again, filters are used to enable specific nodes or nodes located in a particular level to be visualized in a single chart.

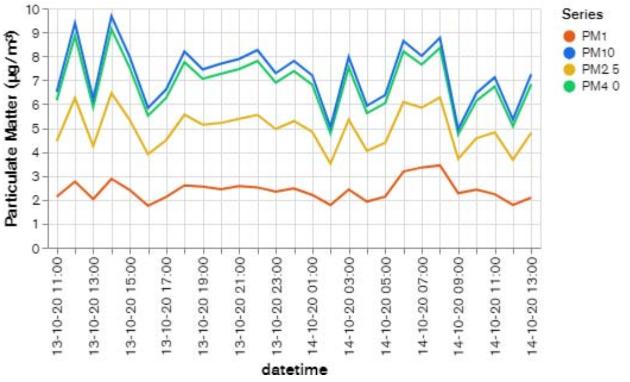


Fig. 7. Daily mean particulate matter graph

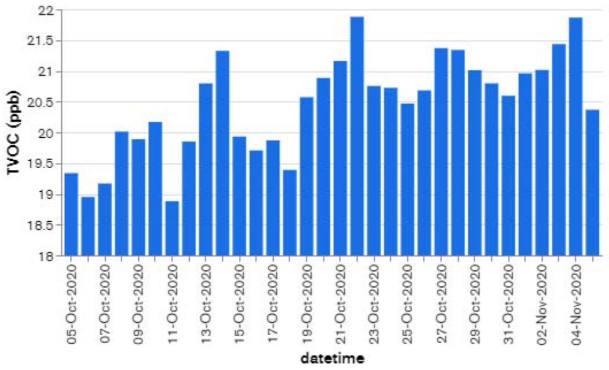


Fig. 8. Monthly mean volatile organic compounds graph

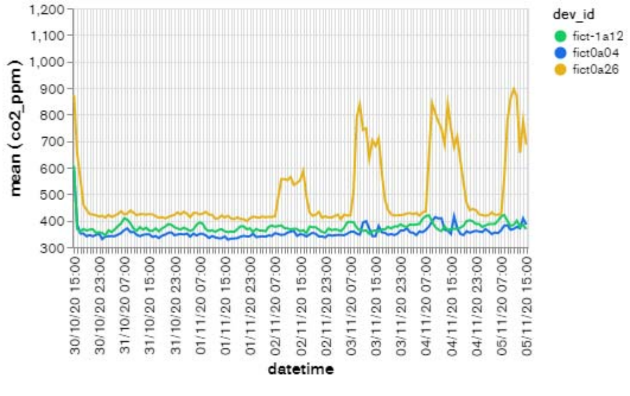


Fig. 9. Weekly mean carbon dioxide from multiple nodes graph

Additionally, Fig. 9 which shows the weekly carbon dioxide graph for three specific nodes, allows us to identify the occupancy time of a particular room throughout the week. The yellow signal shows that this room is occupied during office hours whereas the green and blue signals show us that there was no occupancy during that week in those rooms. Furthermore. the first peak of the yellow signal is lower than the rest of the peaks showing that during that day the number of people occupying that room was lower than usual.

Apart from showing single gases on one graph the system is also able to display multiple gases or atmospheric conditions on one graph. This allows us to analyze the interaction between different gases and different atmospheric conditions. For example, Fig. 10 shows us the interaction of PM10 with changes in temperature and humidity arising from the air conditioning and ventilation systems. This graph can hence give insights on the air filtration performance of the air conditioning and ventilation systems.

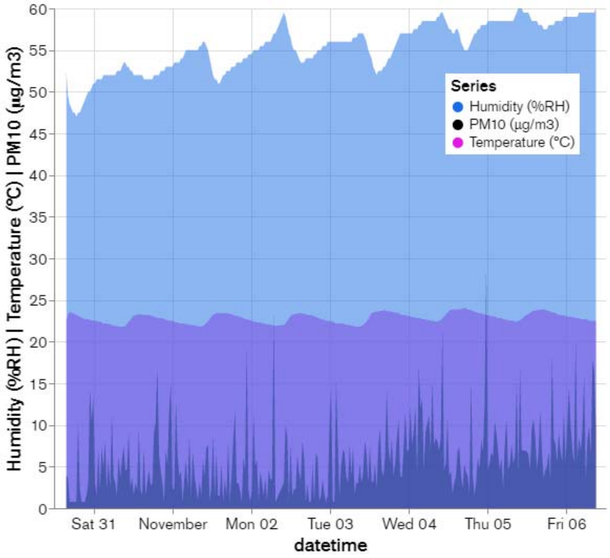


Fig. 10. Weekly mean PM10 in relation to temperature and humidity graph

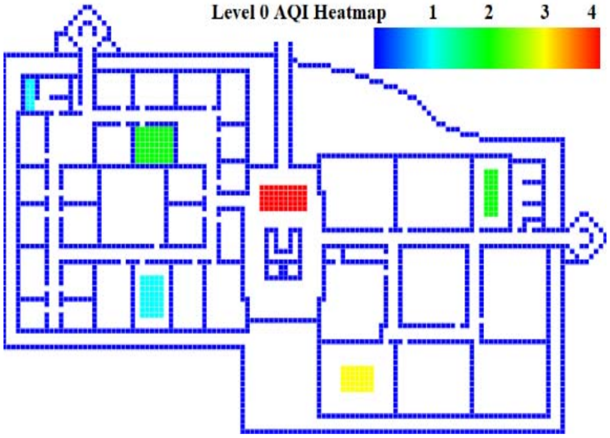


Fig. 11. Level 0 live air quality index heatmap

The system is able to generate both time-based air quality charts and live IAQ data visualizations. Such an example is the AQI heat map, shown in Fig. 11. The AQI heat map gives visual AQI information in relation to the location of the sensor node within the building. Data from these charts can be easily extracted and used to train deep learning models that automatically control the BMS ventilation according to the AQI levels.

The main pollutant chart shown in Fig. 12 is another data analysis result which shows the main pollutants in a particular room over a weekly period. This information enables preventative measures to be taken against specific pollutants such as identifying particulate filters that need to be replaced in specific areas within the ventilation system.

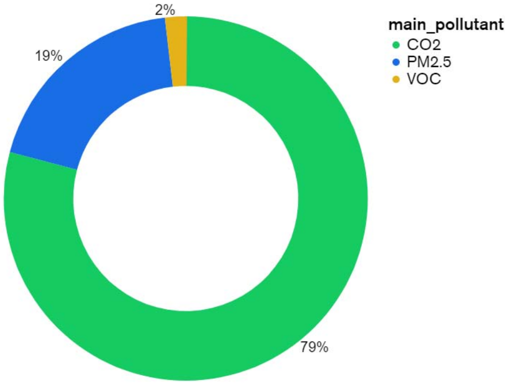


Fig. 12. Main pollutants over a weekly period

V. CONCLUSION

A low cost LoRa-based IoT big data capture and analysis solution for IAQ monitoring was successfully developed. As a proof of concept ten prototype IoT IAQ sensor nodes were deployed alongside one LoRa gateway in an enclosed four floor building made up of a concrete substructure and steel framed upper structure. The IoT nodes can measure temperature, humidity, pressure, carbon dioxide, total volatile organic compounds, and particulate matter. Digital and analogue interfaces on the IoT nodes allow the system to be easily extended to measure additional gases.

The IAQ data captured by the sensor nodes is formatted using Cayenne LPP and transmitted over a LoRa transceiver. A LoRa gateway captures the IoT data and uploads it via a router to a cloud-based LoRaWAN network. An MQTT bridge integrated on the LoRaWAN network publishes the received data. This is captured by an MQTT recorder which formats the data, computes the AQI, TCI, and the corresponding main pollutant. The formatted IAQ data is stored in a document-oriented database. API integrations within the database generate charts and visualizations which are used to perform data analysis the captured IoT big data.

ACKNOWLEDGMENT

This work is part of the research undertaken by the authors in Environmental Sensors for Air Quality (ESAIRQ), which is a PENTA-EUREKA project co-funded by Malta Enterprise.

REFERENCES

[1] S. Bhattacharya, S. Sridevi and R. Pitchiah, "Indoor air quality monitoring using wireless sensor network," *2012 Sixth International Conference on Sensing Technology (ICST)*, Kolkata, 2012, pp. 422-427.

[2] H. Jafari, Xiangfang Li, L. Qian and Y. Chen, "Community based sensing: A test bed for environment air quality monitoring using smartphone paired sensors," *2015 36th IEEE Sarnoff Symposium*, Newark, NJ, 2015, pp. 12-17.

[3] U. Noreen, A. Bounceur and L. Clavier, "A study of LoRa low power and wide area network technology," *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, 2017, pp. 1-6.

[4] U. Raza, P. Kulkarni and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855-873, 2017.

[5] S. Chickerur, A. Goudar and A. Kinnerkar, "Comparison of Relational Database with Document-Oriented Database (MongoDB) for Big Data Applications," *2015 8th International Conference on Advanced Software Engineering & Its Applications (ASEA)*, Jeju, 2015, pp. 41-47.

[6] C. Rattanapoka, S. Chanthakit, A. Chimchai and A. Sookkeaw, "An MQTT-based IoT Cloud Platform with Flow Design by Node-RED," *2019 Research, Invention, and Innovation Congress (RI2C)*, Bangkok, Thailand, 2019, pp. 1-6.

[7] N. Naik, "Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP," *2017 IEEE International Systems Engineering Symposium (ISSE)*, Vienna, 2017, pp. 1-7

[8] S. M. Saad, A. Y. M. Shakaff, A. R. M. Saad, A. M. Yusof, A. M. Andrew, A. Zakaria, and A. H. Adom, “Development of indoor environmental index: Air quality index and thermal comfort index,” *AIP Conferene Proceedings 1808*, 2017, pp. 020043-1–020043-11.