A Cross-lingual Search Engine for Retrieval of Green House Gas Emission Factor

Yachai Limpiyakorn1, Nattapot Nuwong2

1, 2 Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand

Corresponding Author: Email address: yachai.l@chula.ac.th

*Abstract*—*Climate change is a critical global challenge primarily driven by greenhouse gas (GHG) emissions. Managing GHG data effectively is essential for accurate carbon footprint assessments. However, existing information retrieval systems often lack cross-lingual search often lack cross-lingual search capabilities, limiting access to Green House Gas Emission Factor (EF) data across languages, This study present a cross-lingual search engine designed to retrieve EF data effectively, integrating dictionary-based synonym matching and Elasticsearch indexing, The system enables seamless bilingual search functionality, it supports synonym-based search expansion, ensuring relevant result regardless of query language. By implementing an automated pipeline for data retrieval and indexing, the system enhances retrieval accuracy and efficiency thereby aiding organizations in precise carbon footprint calculations.*

Keywords— Cross-Lingual, Information Retrieval, GHGs Emission Factor, Synonym.

# Introduction

The world is currently facing extreme weather conditions, primarily due to the continuous and substantial emissions of greenhouse gases (GHGs) into the atmosphere. These emissions have widespread impacts on ecosystems, climate, biodiversity, and the quality of life at both local and global levels. Consequently, the effective monitoring and management of GHG emissions data is of paramount importance, especially for Thailand. The principal agency responsible for assessing, collecting, compiling, and disseminating GHG-related information is the Thailand Greenhouse Gas Management Organization (TGO), which operates under the Ministry of Natural Resources and Environment. TGO has developed comprehensive databases and report covering a wide range of sectors including industry, energy, agriculture, and communities as well as information for assessing and calculating carbon footprints. Moreover, it determines GHG emissions factors (EF) based on international standards and guidelines provided by the Intergovernmental Panel on Climate Change(IPCC). However, despite the TGO’s efforts to publish GHG emission factor data in various formats on its website (such as online databases and electronic files), there remains a gap the absence of a search system that can comprehensively handle EF queries in a cross-lingual context. In the content provide on TGO’s Website, some EF keywords appear in English, some in Thai, and some in both languages. Additionally, user preferences vary, with some users favoring Thai search terms while others prefer English. These linguistic discrepancies can cause conventional single-language search approaches to yield inaccurate or incomplete result. This research proposes the development of a cross-lingual search program for retrieving of GHG emission factor. The system is designed to support in both Thai and English and to incorporate synonym matching to enhance user accessibility regardless of language. It enables rapid and comprehensive searches, with the search result displaying EF entries across different versions to allow users to select the most appropriate value. This functionality is particularly beneficial for practitioners who need to choose correct EF values when preparing an organization’s carbon footprint report, thereby reducing errors, avoiding redundant work, and ensuring efficient resource utilization.

# Background

1. *UVC (Ultraviolet-C)*

UV radiation is naturally blocked by the ozone layer in the atmosphere. The UV light encompasses three sub-bands: UVC, UVB, and UVA, categorized by its light wavelength spectrum. The UV radiation spectrum's highest energy region is known as UVC. It operates at wavelengths ranging from 200 to 280 nanometers, which are highly effective at disrupting the DNA and RNA of microorganisms. The only way that humans can be exposed to UVC radiation is from an artificial source like a lamp or laser. In 2023, Albertini et al. [3] studied the effectiveness of UVC in reducing microbial load on surfaces and in the air. The results reported significant reductions in healthcare-associated infections (HAIs) when UVC was used as part of a comprehensive disinfection protocol. In 2022. Garg et al. [4] presented work on the efficiency of UVC light to attenuate SARS-CoV-2 transmission, which is timely and relevant. The aerosols containing SARS-CoV-2 were generated using a nebulizer and passed through the UVC disinfection unit. The air exiting the unit was sampled and tested for viral load using the quantitative real-time PCR (qRT-PCR) to determine the reduction in viral RNA. Additionally, the infectivity of the virus was assessed using cell culture techniques. The experimental result reported significant reduction in viral load, achieving up to 99.5% of inactivation rates (inactivation of SARS-CoV-2 at UVC dosages as low as 0.28 mJ/cm²). The simulation suggested that the risk of infection could be substantially reduced up to 90% when the UVC disinfection system was used continuously during occupancy.

1. *LPWAN (Low-Power Wide-Area Network)*

A kind of wireless telecommunication technology, LPWAN, is primarily designed for interconnecting devices that require long-range communication capabilities and battery life sustainability over a broad area [5]. The technology is especially suited for Internet of Things (IoT) applications where devices typically transmit small amount of data over long distances while operating under constrained energy conditions. LPWAN technologies utilize relatively low frequencies within the sub-GHz spectrum, such as 433MHz, 868MHz, and 915MHz, which achieve greater reliability against noise and reducing energy consumption compared to the higher frequency technologies like WiFi and Bluetooth that operate at 2.4GHz. The use of these lower frequencies allows LPWANs to provide the coverage over extensive areas with improved penetration through physical obstacles like buildings and trees.

1. *LoRaWAN (Long Range Wide Area Network)*

Fig. 1 illustrates the architecture where LoRaWAN resides a MAC layer grounded on proprietary LoRa modulation issued by a nonprofit organization known as the LoRa Alliance. Semtech develops the physical layer, or LoRa technology [6]. The LoRaWAN MAC layer protocol is fundamental in managing the communication between end-devices and gateways in a LoRa network. It is responsible for handling the data rate, frequency, and power for each device in the network to optimize both the range and the energy consumption of the devices. This layer operates with different classes of devices: A, B, and C. Each class is designed to meet specific needs in terms of power usage and network behavior [6].

* + *Class A*: The devices consume the lowest power and allow two short-receive windows following each uplink transmission. This class is best suited for applications where the end-device only needs to receive downlink communications after it sends an uplink transmission.
  + *Class B*: In addition to Class A's receive windows, these devices open additional ones at predetermined periods. This is done by using a beacon scheduled by the gateway to synchronize the opening of these windows, allowing for more downlink opportunities.
  + *Class C*: The devices are almost always on and will listen for downlink messages whenever they are not transmitting. This class is suited for applications that require low-latency communication but they consume more power than Class A and B devices.

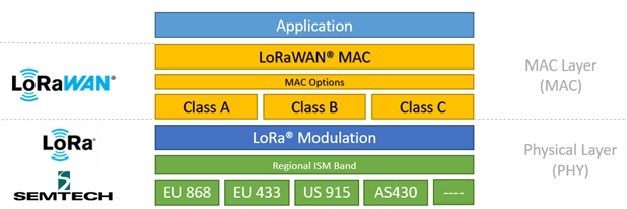


Fig. 1. LoRaWAN (MAC) protocol stack [7].

By integrating these functionalities, the LoRaWAN MAC layer ensures that the network is scalable, secure, and efficient in its power usage, making it suitable for a wide range of IoT applications, particularly those requiring long-range communication with minimal power consumption. In literature, researchers have studied the performance of an IoT application using LoRa Wide Area Network evaluated by widely used of the measures of RSSI and SNR [8, 9, 10, 11, 12].

* + *Received Signal Strength Indicator (RSSI)*− The term "radio strength indicator" (RSSI) refers to a measurement used in wireless networking that ranges from zero to minus 120dBm (decibel-milliwatts), where zero denotes the strongest possible signal and minus 120dBm denotes no detectable signal at all. In this study, RSSI is used for measuring the signal strength between the transmitter (sensor node equipped on the sterilizer), and the receiver (gateway). The stronger RSSI values suggest the robust connection, which is essential for reliable data transmission of the status and functioning of the UVC sterilizer. On the use of LoRaWAN, the higher RSSI values can enhance the reliability of receiving timely notifications and alerts about maintenance needs or system faults. The value of RSSI is calculated as in (1).

*RSSI = –10n log10(d) + I* (1)

where

n = propagation constant, d = distance in meters, and

I = strength of received signal in dBm per meter.

Table 1 summarizes the quality levels of RSSI signal strength denoting reliable data delivery in a LoRaWAN network.

* + *Signal to Noise Ratio (SNR)*− By comparing the desired signal level to the background noise level, SNR provides an indication of the quality of the transmission. The higher SNR values imply clearer signal reception, which directly impacts the accuracy and reliability of the data received from the UVC air sterilizer. High SNR ensures that the data packets sent by the sterilizer, such as operational status and fault indicators, are not corrupted by noise, thus providing accurate and dependable monitoring and control. The general formula of SNR in decibels (dB) is shown as in (2).

*SNR = 10×log10(Psignal) - 10×log10(Pnoise)* (2)

where

Psignal denotes Power of desired signal (in watts or other power units),

Pnoise denotes Power of the noise (same unit of measure as the signal).

Table 2 describes the levels of signal quality of SNR.

* + *Noise*− Noise is an important factor in wireless communication systems, including LoRaWAN, as it directly affects signal quality and the reliability of data transmission. In the context of LoRaWAN, noise refers to any unwanted electromagnetic interference that degrades the quality of the received signal. Typical values range from -90 dBm to -120 dBm, with -120 dBm being very quiet. The noise levels closer to 0 dBm indicate more interference.

Noise is implicitly addressed through discussions on SNR, which is a critical metric used for assessing the robustness of a communication link. A higher SNR implies a cleaner (less noisy) signal, which is crucial for ensuring reliable data transmission over LoRa networks. With higher spreading factors, the concern arises, how LoRa devices achieve better immunity against multipath fading and signal degradation, which are forms of noise [4]. The discussion ties the concept of noise directly to the reliability and effectiveness of the communication channel. Focusing on managing noise through technological strategies like choosing appropriate spreading factors can significantly enhance performance of LoRa-based systems. Regularly measuring noise levels, optimizing node placement, and using appropriate hardware, can mitigate the impact of noise on the network

# METHODOLOGY

Fig. 2 illustrates the system design of an IoT-based UVC air sterilizer that lays the groundwork for both hardware and software development. LoRaWAN is opt to manage the communication between end devices and gateways in a LoRa network. The measures including RSSI, SNR, Noise, are used for evaluating how well the system could monitor and manage the air sterilizers, particularly the communication efficiency and system reliability in a real-world setting. The proposed solution was tested by simulation of typical operational conditions encountered in environments like hospitals or laboratories, where maintaining air quality is crucial.

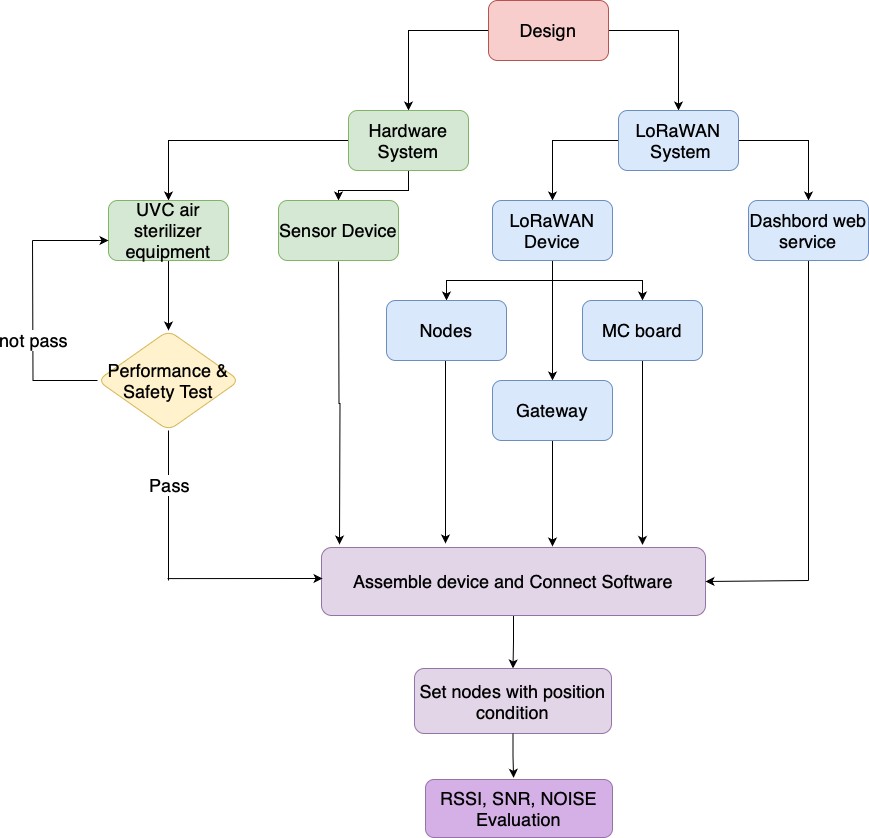


Fig. 2. System design of UVC air sterilizer with LoRaWAN.

As shown in Fig. 3, each component plays a vital role for the overall system functionality. Sensor devices were integrated with LoRaWAN technology to enable continuous monitoring and control capabilities. The implementation started from the sensor which captures critical data passing to the microcontroller for processing and preparing the data for transmission. The integration with LoRaWAN enables seamless data transfer to the gateway, which acts as a crucial intermediary, bridging the local network with the cloud infrastructure via the internet.

TABLE I. Quality rate of signal strength.

|  |  |  |  |
| --- | --- | --- | --- |
| **Quality** | **Signal Strength (dBm)** | **mW** | **Description** |
| Amazing | -30 dBm | 1/1,000th of 1 milliwatt | Max achievable signal strength  Not typical or desirable in the real world |
| Very Strong | -40 dBm | 1/10,000th of 1 milliwatt | Good Level for Wifi connecting |
| Very Strong | -50 dBm | 1/100,000th of 1 milliwatt | Very strong signal with minimal packet loss |
| Very Strong | -60 dBm | 1 millionth of 1 milliwatt | Good Level for LTE and 3G |
| Strong | -70 dBm | 1 ten-millionth of 1 milliwatt | Reliable signal with low to moderate packet loss Good Level for email and web |
| Fair | -80 dBm | 1 hundred-millionth of 1 milliwatt | Minimum signal strength for basic connectivity Packet delivery maybe unreliable |
| Weak | -90 dBm | 1 billionth of 1 milliwatt | Approaching to the noise floor |
| Very Weak | -95 dBm | Noise Floor | Under -90 dBm is no signal |

TABLE II. SNR quality levels.

|  |  |  |
| --- | --- | --- |
| **Quality** | **Range (dB)** | **Description** |
| Very good | >10 dB | Excellent signal quality with minimal interference |
| Good | 6-10 dB | Reliable signal with low interference |
| Fair | 0 - 6 dB | Usable signal, but some interference may occur |
| Poor | -10 - 0 dB | Weak signal with significant interference, communication might be unreliable |
| Very poor | < -10 dB | Very weak signal with high interference, unreliable for communication |

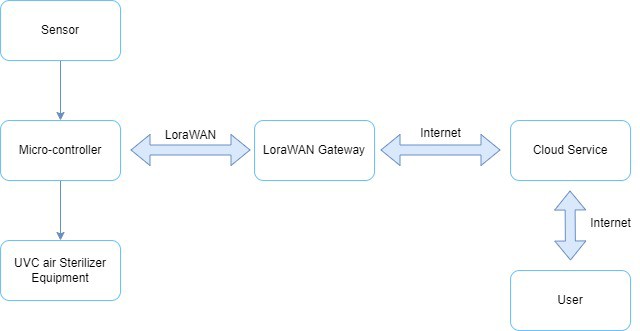


Fig. 3. Design of remote controlling of UVC air sterilizers.

1. *Experiment*

The experiment involved a setup where multiple UVC air sterilizers equipped with sensors were interconnected through LoRaWAN.



Fig. 4. Assembly devices.

These sensors collected various data points, such as operational status and efficiency metrics of the UVC lamps. Each sensor node is linked to a central gateway via LoRaWAN, enabling long-range and low-power communication. The setup aimed to validate the system's capability to function seamlessly in conditions that mimic real-world operational scenarios. The assembly devices are shown in Fig. 4*.*

* + *Location and Configuration*− The nodes were strategically placed in a designated indoor environment to mirror a typical installation in a public building or healthcare setting. The arrangement was designed to test the LoRaWAN network’s ability to handle signals in a space with potential physical and electronic interference, such as walls and other electronic devices that might impact signal quality. In this experiment, one unit of UVC equipment was installed covering the area not exceeding 50 square meters. Several factors need to be considered to ensure the system operates efficiently. The measure of Cubic Feet per Minute (CFM) is used to determine the volume of space and decide how quickly required for exchanging the air in a room. The equation is shown in Fig. 5. Several factors need to be considered to ensure the system operates efficiently, including:

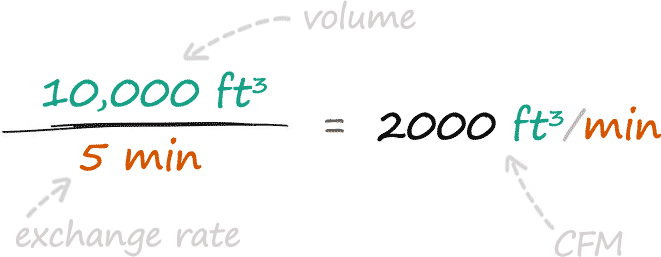


Fig. 5. CFM calculation equation.

− Device Positioning

The device mounted on the wall or ceiling should be at a point that allows the device to cover the entire usable area without any obstructions blocking the signal. The installation should be at a level that can distribute the signal to all important areas in the room effectively.

− Distance from the Gateway

The distance between the node and the gateway should not be too far, as it could cause the signal to drop out easily. The appropriate distance depends on the capabilities of the device and the environmental conditions within the room, including its ability to resist interference.

− Characteristics of Walls and Room Layout

The characteristics of the walls, such as the construction materials used, can affect the propagation of wireless signals. The wall made of concrete or metal may block or reflect wireless signals. Additionally, the layout and arrangement of the room are important as they can affect the distribution and reception of signals between the node and the gateway.

− Pre-installation Testing

Prior to permanent installation, the system should be tested to ensure that the signal from all devices can connect well with the gateway and that the system is stable. Fig. 6 illustrates the testing scenario of room size 5x8 meter containing eight UVC Lora nodes and a LoRa gateway.

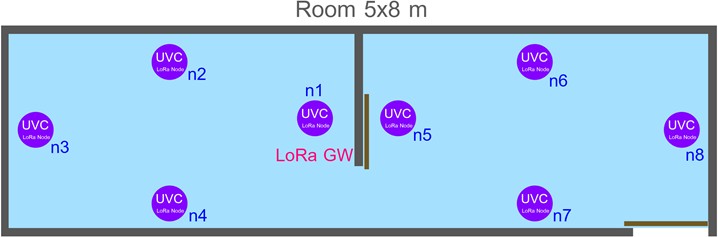


Fig. 6. Layout of LoRa gateway and nodes in experiment room.

1. *Evaluation*

The outcomes of the experiment provided insights into the practical deployment of IoT technologies in managing air quality control systems. The collected data helped in understanding the interaction between device placement and signal efficacy, which is crucial for deploying IoT solutions in complex environments. The measurement data gathered from each node, RSSI, SNR, and Noise, were used to compare with the threshold of each parameter to ensure that the system and each node function properly without any effect to decrease communication efficiency as well as to identify any potential drop in performance due to environmental factors.

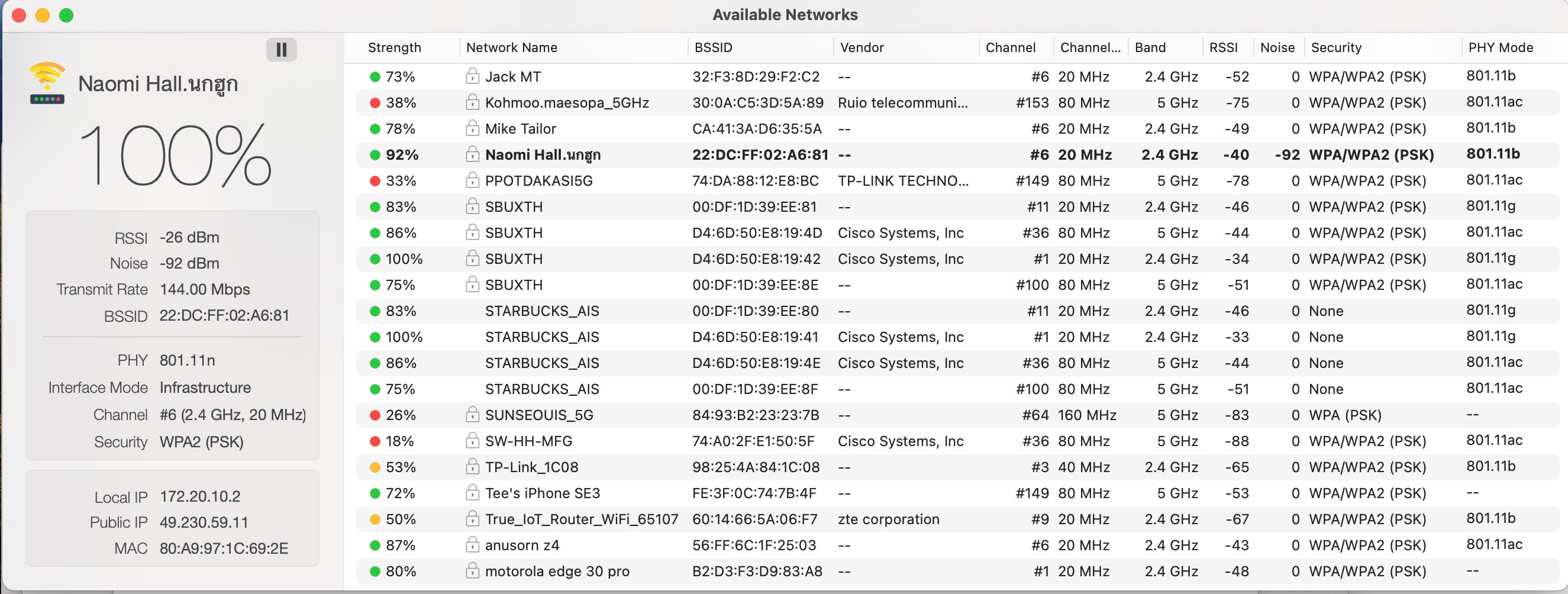


Fig. 7. Example information resulting from using WiFi signal strength analyzer application.

Fig. 7 reports the measurement data of RSSI and Noise by coding program on MC board or using application. RSSI is a measure of the power level that an individual device receives from the wireless access point or router. The measurement data gathered from the network “Naomi Hall” reported RSSI of -40 dbM (very strong signal) where -30 dBm (excellent) and -90 dBm (very weak). The value closer to zero denotes the stronger signal. The Noise level was reported -92 dBm, where typical values range from -90 to -120 dBm. Contrary to RSSI, the value of Noise closer to zero indicates more interference. The value of SNR was computed from the difference between the signal strength and the noise level. Here, the calculated value of SNR equals 52 dB considered excellent signal quality with minimal interference.

1. *Measurement Data*

Example measurement data of node2, 3, 4 with timestamp sending between MC board and nodes were collected. The graphs that visualize the measurement data of signal strength (RSSI), signal-to-noise ratio (SNR), and the presence of any noise interference affecting communication and visualized are shown in Fig. 8, 9, 10, respectively. These measurement data were useful for assessing the network's robustness and the reliability of data transmission across the system.

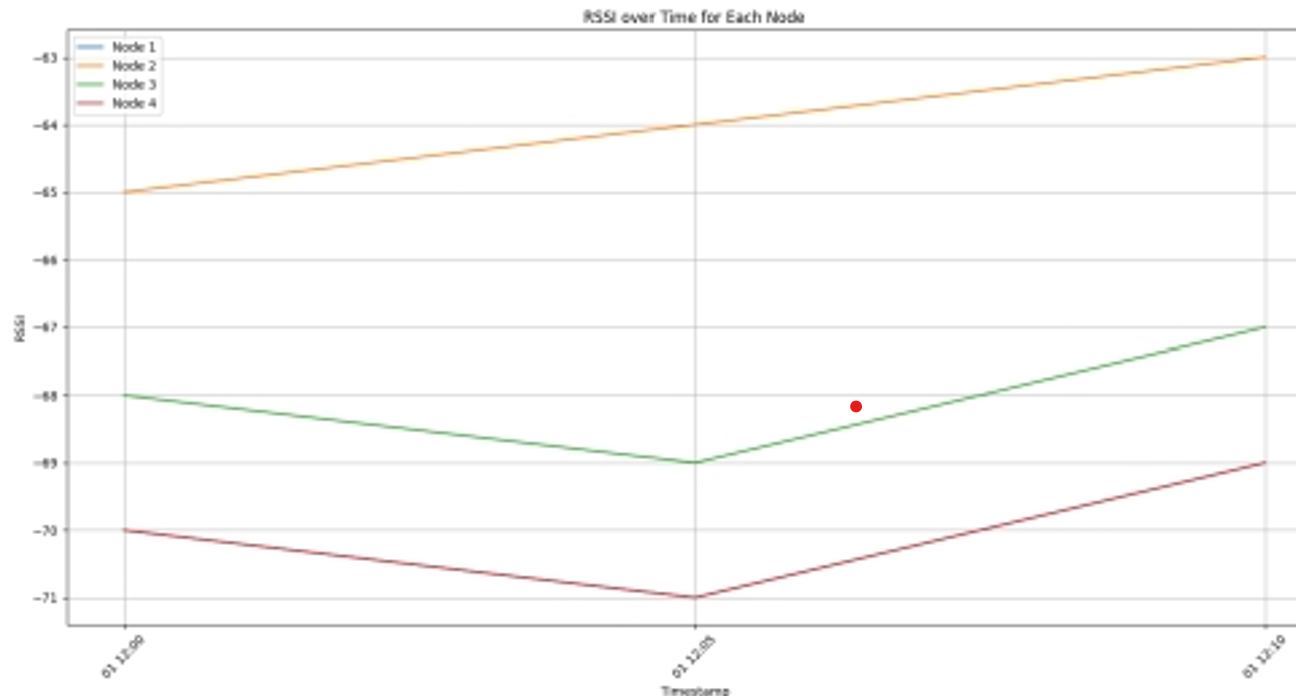


Fig. 8 RSSI over time.

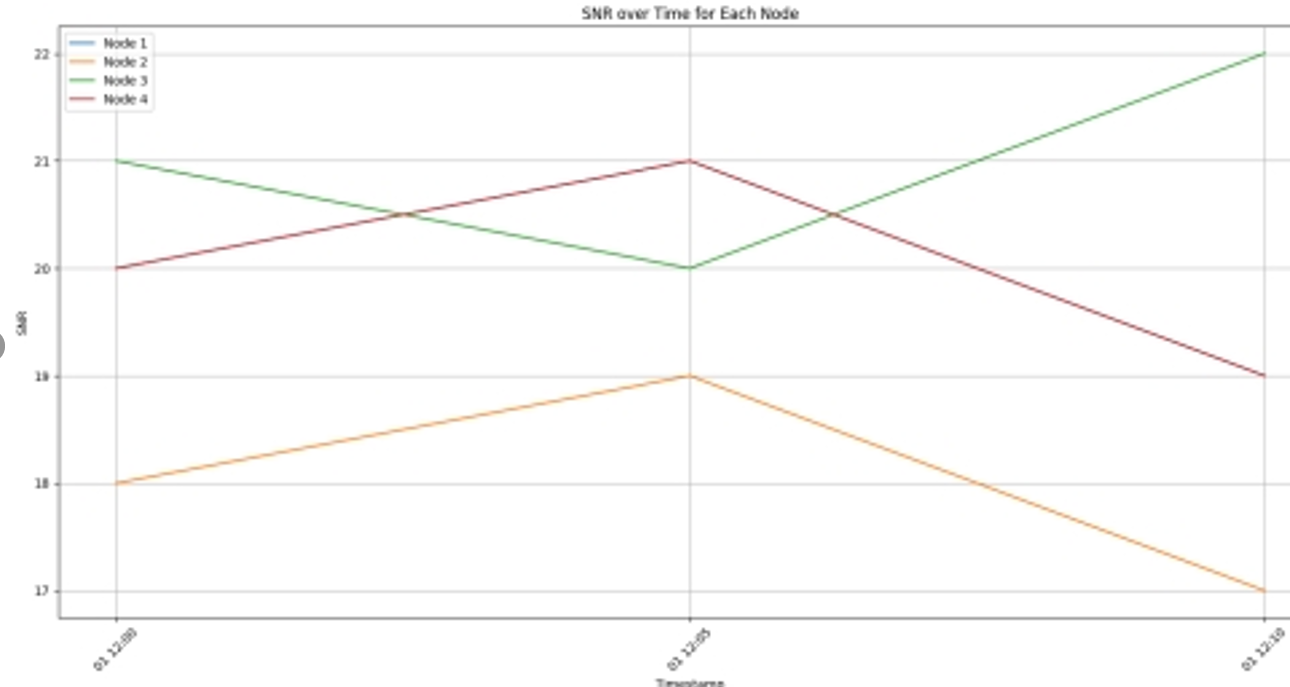


Fig. 9 SNR over time.

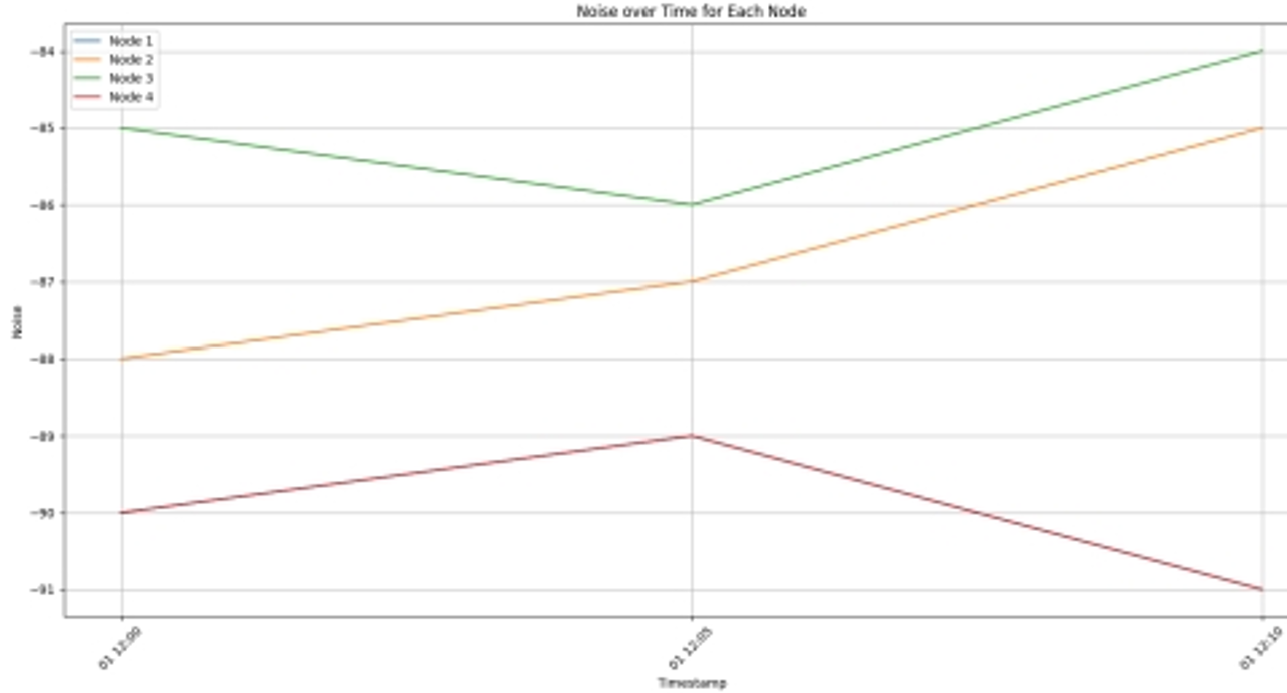


Fig. 10 Noise over time.

1. *Control and Notification*

The Control and Notification module is implemented with ChirpStack LoRaWAN for real-time monitoring the UVC air sterilizers. The functionality includes:

* + *Event Tracking:* Each event in the system is logged with a timestamp, indicating activities such as data transmissions and system checks. For instance, the system records at 2024-05-20 15:25:17 reporting successful uplinks, suggesting robust communication between the devices and the network server.
  + *Data Transmission:* The data columns indicate transmission of multiple data points, which likely include operational status, error codes, and possibly real-time UVC intensity levels. The data are crucial for ensuring the devices are operating safely and effectively, without exposing occupants to harmful UVC levels*.*
  + *Real-time Alerts:* Immediate notifications will be sent for warning any event of deviations from normal operating conditions, such as a drop in UVC intensity or system malfunctions. This proactive notification helps maintaining continuous sterilization effectiveness and immediate troubleshooting.
  + *System Logs and Updates:* The continuous logging of events ensures that all changes and incidents are recorded, allowing for detailed analysis and future improvements.

Fig. 11 illustrates an excerpt from the dashboard of control and notification module.

# CONCLUSION

This research underscores the potential of IoT technologies like LoRaWAN to revolutionize building management systems contributing to public health safety. The methodology demonstrates the effectiveness of using LoRaWAN cloud- based solution for remote monitoring and controlling of UVC air sterilizers by systematically addressing each phase, from design to evaluation. The proper system configuration, including the placement of LoRaWAN nodes and the tuning of signal-related variables such as RSSI, SNR, and Noise, potentially influence the performance and reliability of the communication links. The findings reported significant improvements in operational efficiencies and proactive maintenance capabilities that would lead to safer indoor environments.

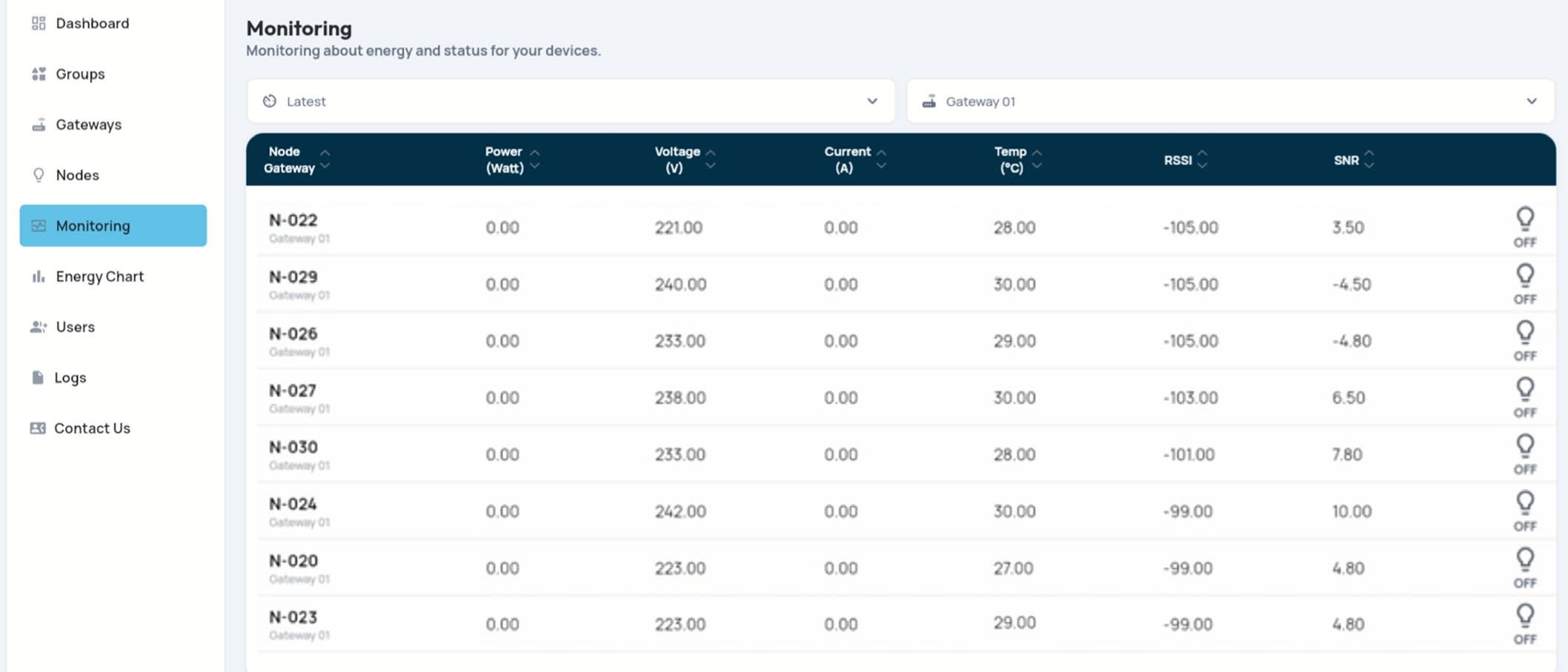


Fig. 11. Example dashboard of control and notification module.

The experiment successfully demonstrated the implementation and efficiency of a LoRaWAN-based control and notification system for UVC air sterilization in critical environments like hospitals and laboratories. Integrating LoRaWAN, a low-power, long-range communication technology, would enhance the ability to remotely monitor and manage UVC air sterilizers, ensuring sustainable air quality and reducing the risk of airborne infections.

Future direction would be optimizing the system architecture and extending the system's applicability to other types of environmental monitoring, in addition to exploring advanced data analytics to more accurately predict maintenance needs. For example, machine learning algorithms could provide deeper insights and predictive maintenance capabilities, further improving system performance and reliability.

REFERENCES

1. M. R. Bagsiraj, S. R. Narwade, and K. S. Pirjade, “A Review on Air Sterilizer for Purification of Air Using UV Light and Filters,” Asian Journal of Science and Applied Technology, vol.10, issue 2, pages 34- 36, 2021.
2. M. Luvisotto, F. Tramarin, L. Vangelista, and S. Vitturi. “On the Use of LoRaWAN for Indoor Industrial IoT Applications,” Wireless Communications and Mobile Computing, Hindawi, 2018.

[3] R. Albertini, M. E. Colucci, and A. Coluccia, “An overview on the use of ultraviolet radiation to disinfect air and surfaces,” Acta Biomed; vol. 94, Supplement 3: e2023165, 2023.

1. H. Garg, R. P. Ringe, and S. Das, “UVC-Based Air Disinfection Systems for Rapid Inactivation of SARS-CoV-2 Present in the Air,” Pathogens, vol. 12, issue 3, page 419, 2023.
2. A. A. Rahman, F. K. Zaman, and S. A. Abdullah. “Performance Analysis of LPWAN Using LoRa Technology for IoT Application,” International Journal of Engineering & Technology, vol. 7, issue 4.11, pages 252-256, 2018.
3. J. Haxhibeqiri, A. Karaagac, F. V. Abeele, W. Joseph, I. Moerman, and J. Hoebeke, “LoRa Indoor Coverage and Performance in an Industrial Environment: Case Study,” in ETFA, Limassol, Cyprus, 2017.
4. Wi-Fi Vitae, “LoRaWAN overview,” [Online] Available: https://wifivitae.com/2023/02/07/lorawan-overview-2/. [Accessed: 22 May 2023].
5. S. S. Ahmed and B. I. Saeed, “LoRa: A Proposed Connectivity Technology for Internet of Things Applications in the Kurdistan Region of Iraq,” Kurdistan Journal of Applied Research, vol. 6, issue 2, pp. 20- 34, 2021.
6. M. A. Mojamed, “On the Use of LoRaWAN for Mobile Internet of Things: The Impact of Mobility,” Applied System Innovation, vol. 5, issue 1, 2022.
7. P. Gotthard and K. Minolta, “Low-Cost Car Park Localization Using RSSI in Supervised LoRa Mesh Networks,” in WPNC, Bremen, Germany, 2018.
8. L. Aarif, M. Tabaa, and H. Hachimi, “Experimental test and performance of RSSI-based indoor localization in LoRa Networks,” Procedia Computer Science, vol. 203, pp. 420-425, 2022.
9. M. N. Bin, C. Kamarudin, and A. B. Ayob, “Review Of LoRaWAN: Performance, Key Issues, And Future Perspectives,” Jurnal Kejuruteraan, vol. 36, issue 2, 2024.