Automatic Parameter Adjustment for Hybrid WiFi and BLE-based Congestion Measurement

Akira Tokuda*, Yutaka Arakawa*, Shigeru Takano*, Shigemi Ishida†

* *Kyushu University*, Fukuoka, Japan, Email: {tokuda.akira, yutaka}@arakawa-lab.com, takano.shigeru.019@m.kyushu-u.ac.jp.
* *Future University Hakodate*, Hokkaido, Japan, Email: ish@fun.ac.jp.

***Abstract*—The recent COVID-19 pandemic has made it extremely important to avoid crowded environments. In light of this, we are developing a congestion measurement system utilizing IoT technology, in which the congestion state of a given location is estimated by counting the number of WiFi Probe Request messages and BLE exposure notification messages sent from smartphones and laptops. Our congestion measurement system, however, suffers from an unstable measurement problem because the number of WiFi and BLE messages is highly dependent on the environment in which our system is deployed. This paper presents a congestion measurement system employing an automatic parameter adjuster that accurately estimates the number of people in various target locations. The parameters are automatically adjusted based on the seating capacity and area size of the target environment, as well as information gathered from a sensor which is designed to collect and analyze WiFi and BLE messages collected in a university cafeteria.**

I. Introduction

Demand for congestion information has been increasing in recent years, with traffic congestion information in especially high demand due to the continued spread of COVID-19. Along with road traffic congestion information, people are interested in the congestion information of many public places such as restaurants and cafés, and public transportation like buses and trains. Such congestion information is recently provided by map services such as Google Maps, and is being sold by mobile phone carrier companies and location-based service providers.

We have also installed congestion sensors at bus stops and cafeterias on our university campus since June last year to collect congestion information. These congestion sensors make use of congestion measurement technologies based on WiFi and Bluetooth Low Energy (BLE) wireless signals, which are reported in [1]–[6]. The sensors estimate congestion level by counting the number of WiFi Probe Request (WPR) messages sent from WiFi devices when they connect to a nearby access point. We also make use of the number of BLE exposure notification messages sent from COVID-19 contact confirmation applications, which are becoming prevalent nowadays as a countermeasure against the spread of COVID-19. Compared to congestion measurement using cameras [7]–[11] and mobile-phone base stations [12], the WiFi and BLE-based approach presents substantial advantages in terms of both privacy and data processing-related computational cost.

There are, however, certain drawbacks to WiFi and BLE-based congestion measurement. The latest generation of smartphones has started randomizing the MAC address in WPR messages for privacy protection. Counting the number of observed MAC addresses returns a greater number of smartphone users because a single smartphone transmits probe requests with different MAC addresses for each surrounding SSID. While the BLE-based method counts the exposure notification messages used by the COVID-19 contact verification applications, not all smartphone users install the COVID-19 contact tracing application. In addition to the problem of unreliable people counting, we need to define what is meant by *congestion* because congestion is dependent on many parameters such as the size of target environments. For example, a small room can be considered congested when there are 10 people inside it, while a spacious room can require the presence of at least 100 people before being considered congested.

In this paper, we present an automatic parameter adjustment method for a WiFi and BLE-based congestion measurement system. We discuss the following three points by comparing the actual number of people with measurement results derived using a WiFi/BLE-based congestion measurement system installed on our university campus.

- WiFi-based counting parameters: received signal strength (RSSI) thresholds, correction functions
- BLE-based counting parameters: RSSI thresholds, correction functions
- Definition of congestion level based on the number of observed WiFi and BLE messages and the area size of the target location

The WiFi/BLE-based congestion sensors were installed in two cafeterias on our campus and collected the number of WiFi and BLE messages from 11am to 2pm every weekday. At the same time, we visually counted the actual number of people in each cafeteria every 10 minutes. We designed a congestion level estimation algorithm by comparing the collected data with the actual number of people.

II. Related Work

Research on congestion measurement has been conducted since before the COVID-19 pandemic and has become of great

TABLE I: Comparison chart of congestion measurement technologies

Technology	Technology Overview	Advantages	Disadvantages
Cellular Network	Smartphone location information such as DoCoMo Mobile Spatial Statistics	• No additional equipment	• Wide coverage
Camera	Detects and counts the number of people by object recognition	• High precision	• Only visible areas are analyzed • Restrictions on the location of the camera • GPU processing required
WiFi	Counts Probe Requests	• Low cost, currently on the market	• Medium precision
	Estimated from signal propagation	• High precision	• Specific transmitter/receiver required
BLE	Counts exposure notifications	• iOS and Android compliant	• Installation rate of notification application unknown and changing
Microphone	Measures noise level	• Device-free (effective for children, elderly, etc.)	• Cannot use in quiet environments such as classrooms
CO_2	Measures CO_2 concentration level		• High latency

interest recently. Table I summarizes existing congestion measurement technologies and their characteristics, advantages, and disadvantages.

Population estimation based on cellular network connection information [12] is a simple yet effective way to derive congestion information covering a nationwide area with no additional equipment. The cellular network-based approach, however, cannot provide congestion information for small facilities such as a cafeteria because the cellular network counts the number of people in larger 250m or 500m mesh areas.

Cameras are another type of widely-installed congestion sensor. Arai et al. presented a privacy-preserving congestion measurement method for railway stations utilizing image analysis while avoiding individual human detection [13]. We also presented a selection of camera-based congestion measurement methods for bus stops and shopping streets [7]–[11]. Camera-based approaches, however, have limited coverage and are only capable of counting people in the camera's field of view.

To measure congestion levels in non line-of-sight situations, congestion measurement methods based on wireless technologies such as WiFi and BLE have been proposed. WiFi packet sensors [1]–[6] are an old yet effective method to estimate congestion levels. WiFi packet sensors count the number of WPR messages, which are sent when WiFi devices establish a connection with a nearby access point. There are existing WiFi packet sensors already on the market such as the Ad Intec, Inc. AIBeacon[1]. More recent technology estimates congestion levels based on WiFi signal propagation changes [14]. WiFi-based approaches, however, are likely to over-count the number of people because WiFi devices are prevalent nowadays. MAC address randomization, recently utilized for privacy protection, also makes it difficult to accurately count the number of people.

BLE-based congestion measurement methods count the number of BLE messages in the same way as the WiFi packet sensors. We can easily count the number of exposure notification messages, which are periodically sent from COCOA[2], a COVID-19 contact tracing application. Unfortunately, the estimated installation rate of COCOA is low.

Microphone-based and CO_2 sensor-based approaches have also been proposed. Microphone-based methods cannot be used in quiet situations where people are not speaking, such as waiting for, or riding a bus. CO_2 sensor-based approaches are only applicable in closed spaces and suffer from high latency [15].

We developed congestion sensors utilizing a combined WiFi and BLE-based approach, which count the number of WPR and BLE exposure notification messages. The congestion sensors have been installed at bus stops and cafeterias to collect congestion information.

The collected congestion information, however, was found to be unreliable when compared to the actual number of people in each sensor location. We found that the area size, as well as the nature of the surrounding environments highly affects the number of observed WiFi/BLE messages, which resulted in the discrepancies between the estimated and actual number of people. To accurately estimate congestion levels, we need to properly adjust packet filtering parameters such as minimum signal strength.

III. Hybrid Congestion Level Estimation

In our proposed system, we observe the number of WPR messages sent from devices such as smartphones in the vicinity of WiFi access points, and the number of BLE exposure notification messages sent from the COCOA smartphone application. We propose a congestion evaluation method which makes use of the WiFi and BLE transmission information to estimate the congestion level of a target location with high accuracy.

A. Filtering Private MAC Addresses

Recently, several companies have faced criticism for collecting the unique MAC addresses of their users' devices

[1]https://www.aibeacon.jp/

[2]https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html

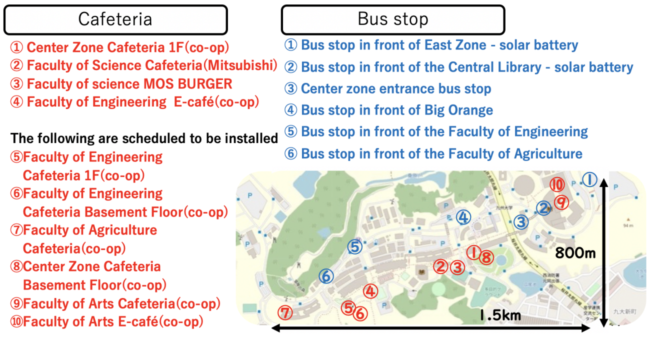


Fig. 1: Map of sensor locations on campus

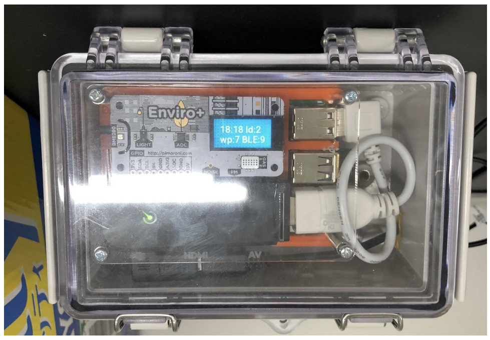


Fig. 2: Installed congestion sensor

and tracking their activities. In response to this, as a privacy measure the latest smartphone operating systems (iOS14 and later, Android 10 and later) randomize MAC addresses. This address is called private MAC (PMAC), as by transmitting using random MAC addresses, they cause WiFi packet sensors to count more devices than are actually present in a given location. This is resolved by counting the MAC and PMAC addresses separately, enabling the process to be performed using only the PMAC addresses. By performing regression between the true value measured visually and measured value of the PMAC, we can obtain a coefficient which enables us to back-calculate the true value from the measured value. In this way, the randomization of MAC addresses is taken into account, which is not the case with conventional WiFi packet sensors. Previous work has proposed methods to eliminate the effect of MAC address randomization through observation of the timings of the network scans with an off the-shelf WiFi interface [16]. It presents a straightforward way to identify PMACs because the second character in a PMAC address is always a 2, 6, A, or E. We make use of this property to eliminate the PMACs in our paper.

B. WiFi/BLE-Based Congestion Sensors

Our proposed system consists of sensor terminals installed at the locations shown in Figure 1, a cloud system that aggregates the information and calculates the degree of congestion, and a mobile website that visualizes the information. The congestion sensor installed is shown in Figure 2. The sensor is based on the Raspberry Pi 3 and uses the environmental sensor board Enviro+[3] in conjunction with an external USB WiFi module and a USB LTE adapter. Both the external USB WiFi and onboard BLE module are operated in monitor mode to capture packets in the surrounding area, and the PMAC filtering and minute-by-minute quantification processes are performed in the sensor terminal, with the obtained numbers sent to the cloud system through the LTE connection. The terminal program runs as a script in Python and is set to start automatically when the sensor power is turned on, and the actual measured values are displayed on the LCD monitor every minute for operation monitoring purposes.

[3]https://shop.pimoroni.com/products/enviro?variant=31155658489939

Our proposed system defines three different levels of congestion. The reason for this is that, depending on the installation environment, the size of the target location may vary greatly. For example, if the number of packets received is equivalent to 30 people, then a cafeteria with 20 seats can be considered crowded, while a cafeteria with 200 seats can be considered empty. For this reason, we thought it would be easier for users to understand if the parameters were automatically adjusted according to the installation location and always shown on the same three levels.

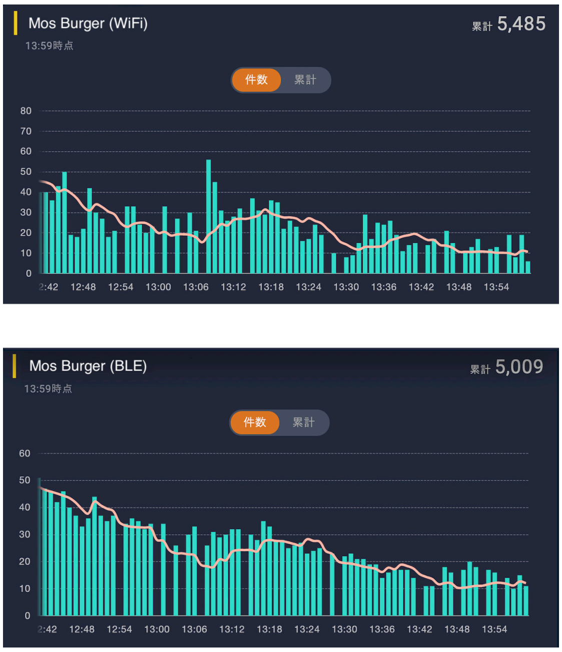
The cloud system estimates the congestion level every 10 minutes based on the number of WPRs and exposure notifications sent from each sensor terminal every minute. Figure 3a and Figure 3b show the sensor monitoring screen displayed to the administrator. The horizontal axis is time, and the vertical axis displays the number of notifications received per minute. Large cafeterias have larger overall count numbers than small cafeterias, and in both cases the number of WPRs tends to be larger than the number of BLE transmissions.

C. Automatic Parameter Adjustment

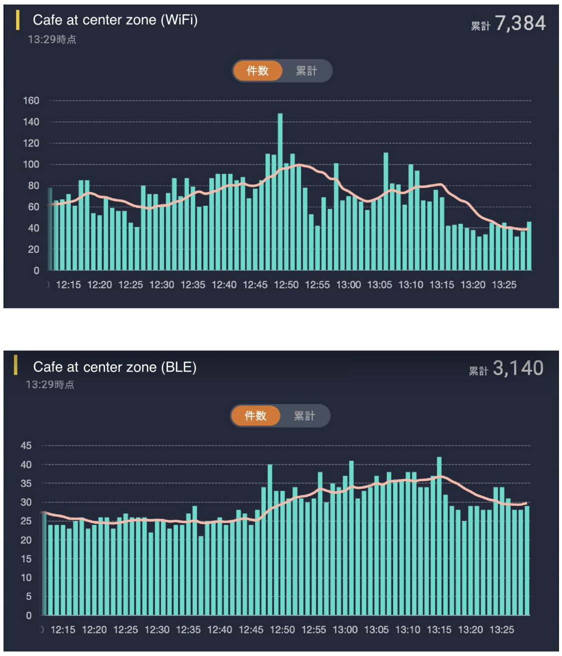
To determine how to design our parameter adjustment method, we visually counted the number of people in the cafeteria while simultaneously recording the values obtained by the installed sensors. Table II and Table III show the measured values and sensor data for a given day. The values are averaged over 10 minutes. The RSSI (Received Signal Strength Indicator)[dBm] records the received signal strength when it passes a pre-determined threshold value. As shown in Table IV, there is a strong correlation between the number of received WiFi and BLE transmissions and the actual number of people at the target location.

However, we found some issues when comparing the values measured by the sensors to the actual values. The sensor acquires all WiFi signals sent from multiple terminals such as smartphones and PCs owned by cafeteria users and devices near the location where the sensor was installed, making the use the acquired values as congestion values challenging.

As for the BLE signals, people who did not have the application installed were not counted, making it difficult to use BLE information to determine the correct number of people.



(a) In small cafeterias



(b) In large cafeterias

Fig. 3: Number of WPR and BLE

TABLE II: Number of WPRs and BLE Exposure Notifications in small cafeterias

time	WiFi Probe Request				BLE (Exposure Notification)				true value
	All	RSSI>-80	RSSI>-70	RSSI>-60	All	RSSI>-80	RSSI>-70	RSSI>-60	
11:00	10.1	10.1	7.7	2.5	8.4	1.9	0.4	0.3	5
11:20	9.8	9.8	7.5	1.8	10.5	2.5	1	0	6
11:40	23.9	23.9	20.3	9	17.1	5.5	2.4	1.4	8
12:00	41	41	36.1	13	27.1	11.6	3.7	2.3	11
12:20	36.1	36	29.6	8.7	49.8	12.6	7	1.3	15
12:40	36.3	36.2	30.9	14.1	42.5	19.8	8.5	1.5	22
13:00	34.5	34.4	29.9	10	29.3	10	3.5	1.4	12
13:20	19.5	19.5	16.6	4	24.7	8.1	1.8	0	7
13:40	16.8	16.8	13.5	4.4	15.4	5.4	2.9	1.6	8
14:00	9.7	9.7	7.1	1	8.6	3	1.3	0.2	5

When we examined the size of the cafeterias used as testing locations, we found that large cafeterias could not detect signals from the devices of users located far away from the sensors, in which case the measured values were often lower than the actual values, while small cafeterias picked up signals from the devices of students passing outside the cafeteria, and the measured values were often higher than the actual values.

We devise a solution to this problem. First, we add any MAC addresses that have been detected and acquired multiple times over a long time period to a blacklist. These signals correspond to devices permanently installed near the sensors and the blacklist eliminates their effect on congestion estimation. In addition, the device manufacturer's information can be used to determine whether the device is an installed device or not. If the manufacturer mainly deals with WiFi routers, we can determine that it is most likely an installed device.

Second, we increase the number of sensors used in large cafeterias, and decrease the number of sensors used in small cafeterias. In addition, in small cafeterias we set a signal strength threshold value, and weak signals who fall below this threshold value are not counted, as they are presumed to be emitted by passers-by. Also, when multiple sensors are used in a large cafeteria to compensate for the lack of radio coverage, the MAC address information acquired by the sensors can be shared among them to prevent duplicate counts within the same cafeteria.

We design an algorithm capable of implementing our proposed solutions. We begin by obtaining the capacity X of the cafeteria whose congestion level is to be monitored. This number is set according to the cafeteria. Next, we set a threshold value Z and count only the signals that exceed this value. An initial value of Z=-80, was used as the minimum threshold in Table II and Table III.

Next, we compare the received signal strength with the threshold value Z, and if it is lower, the signal is not counted. In addition, as a measure to prevent counting passers-by, we do not count a WiFi signal the first time it is observed, we count it only after it has been observed twice.

TABLE III: Number of WPRs and BLE Exposure Notifications in large cafeterias

time	WiFi Probe Request				BLE (Exposure Notification)				true value
	All	RSSI>-80	RSSI>-70	RSSI>-60	All	RSSI>-80	RSSI>-70	RSSI>-60	
11:00	18.7	17.8	12.9	6.2	5.9	3.5	2.4	0.8	17
11:20	26	25	16.3	7	10.6	8.6	4.2	1.4	24
11:40	46.9	45.9	33.1	13.3	15.4	13.3	7.6	2.4	46
12:00	55.4	54.2	39.6	14.4	26.2	22.2	12.9	2.5	66
12:20	60.2	59.8	50.2	19.9	25.2	18.8	11.3	2.7	87
12:40	95.2	93.3	73.5	23.2	28.9	25.6	16	3.4	81
13:00	73.3	73	64.4	22.9	35.8	34.4	22.3	4	94
13:20	39.1	38.7	32.5	13.6	29.8	27.2	18.2	4.8	68
13:40	25.4	24.6	18.4	8.4	20.1	17	12	2.8	40
14:00	31.3	30.4	22.3	8.2	13.4	10.9	8.1	1.9	33

TABLE IV: Correlation coefficients between measured values and actual values

cafeteria	WiFi Probe Request				BLE (Exposure Notification)			
	All	RSSI>-80	RSSI>-70	RSSI>-60	All	RSSI>-80	RSSI>-70	RSSI>-60
small cafeteria	0.8091	0.8090	0.8131	0.8962	0.8122	0.8786	0.9408	0.6999
large cafeteria	0.9054	0.9114	0.9374	0.9482	0.9325	0.8964	0.8286	0.7883

Since BLE transmits signals more frequently than WiFi, we only count BLE signals after they have been observed ten times.

These processes are performed for one minute, and the summed count Y is recorded. Since this process eliminates factors that adversely affect congestion estimation, there is not expected to be a significant difference between the estimated and actual measurements at this point.

The degree of congestion is calculated by comparing this count Y with the number of people the cafeteria can accommodate X. If the measured value Y is larger than the number of seats X, we judge that the threshold value Z was set incorrectly and increase it by a value of 10.

This was judged based on the number of WPRs and the number of BLE exposure notifications: for either transmission type, if $Y \geqq 0.8X$, the target location was judged to be $Congested$, if $Y \geqq 0.3X$, the target location was judged to be $Moderate$, and if $0.3X > Y$, the target location was judged to be $Vacant$. Using data from both WiFi and BLE transmissions enabled us to improve congestion estimation accuracy.

A flowchart of the algorithm based on our solution is shown in Figure 4.

IV. Discussion

Our system showed that there is a high correlation between the actual number of people counted and the values obtained by the sensors, and confirmed that it is possible to estimate the degree of crowding from the sensor values. In the future, we would like to be able to estimate the number of people automatically from the sensor values and compare this number of people with the capacity of the target location to quickly and simply visualize the degree of congestion. In the case of a university cafeteria, we would like to make it possible for students to easily view information regarding the level of congestion. This information could be used to suggest to students to stagger their eating times when the cafeteria is crowded, or to encourage them to use a different, less congested, cafeteria. It is important to note that our method adjusts the RSSI threshold only when the estimated number of people is clearly larger than the room capacity. This approach might overestimate the congestion level, and requires further fine-tuning.

In this paper, we do not cover the installation of our system at bus stops. One of the issues with installing sensors at bus stops is that, due to their proximity to the road, they are prone to detect the signals from the smartphones of the passengers and drivers of passing vehicles. Algorithms that can be adapted to each environment are needed. Our next step is to study the effectiveness of our proposed algorithm under different conditions, not only in cafeterias.

V. Conclusion

At Kyushu University, the ban on some face-to-face lectures has been lifted and students have resumed commuting to school, and the number of users of the cafeteria is expected to increase further. At the same time, new, more infectious COVID-19 variants are spreading, and infection control is becoming even more important. In the future, we would like to further improve our congestion level estimation system so that student users can easily check the congestion status of a wide variety of locations on campus.

Acknowledgment

This work was partially supported by a Center of Innovation (COI) program of the Japan Science and Technology Agency (JST) (JPMJCE1318), and the Cabinet Office (CAO), Cross-ministerial Strategic Innovation Promotion Program (SIP), “An intelligent knowledge processing infrastructure, integrating physical and virtual domains” (funding agency: NEDO) (JPNP18014).

References

[1] T. Jozen, S. MIKAMIYAMA, Y. Tsujimoto *et al.*, “A performance experiment of the wi-fi packet sensor for traffic flow sensing,” *The 49th Spring Meeting of the Civil Engineering Research Conference of the Civil Engineering Society*, no. 49, 2014.

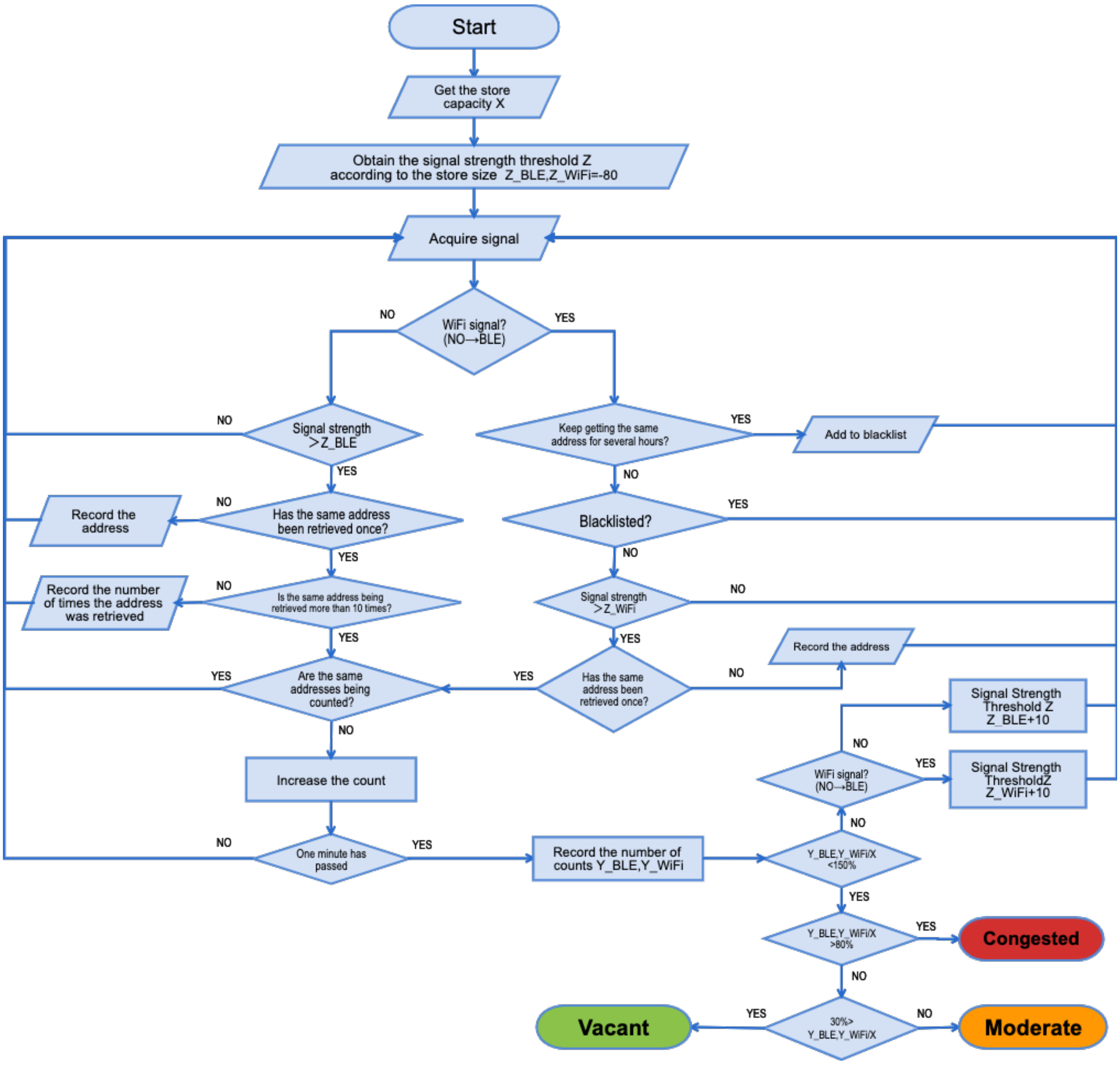


Fig. 4: Flowchart of automatic parameter adjustment

[2] M. MOCHIZUKI, T. Jozen, J. NISHIDA *et al.*, "Construction of an anonymous human flow analysis system using wi-fi packet sensors," *UBI*, vol. 2014, no. 45, pp. 1–8, 2014.

[3] T. Morimoto, Y. Tsujimoto, S. Shirahama *et al.*, "Human flow analysis and visualization using Wi-Fi packet sensors," in *DEIM Forum*, 2015.

[4] K. Ichii, S. Terabe, H. Yaginuma *et al.*, "Understanding travel behavior among tourism attractions by wi-fi scanner," *Journal of Civil Engineering and Planning*, vol. 57, pp. 01–16, 2018.

[5] A. Shimada, K. Oka, M. Igarashi *et al.*, "Congestion analysis across locations based on wi-fi signal sensing," in *6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, Jan. 2018, pp. 204–221.

[6] S. Nakata, A. Okamoto, and M. Horikawa, "Congestion degree estimation using probe request in wireless lan," *DPS*, vol. 2018, no. 1, pp. 1–7, 2018.

[7] S. Takano, M. Hori, Y. Arakawa *et al.*, "Towards ict based mobility support system with in the covid-19 era," in *The 18th ACM Conference on Embedded Networked Sensor Systems (SenSys 2020)*, 2020.

[8] R. Takahashi, K. Hayashi, Y. Mitsukude *et al.*, "Bus stop congestion visualization system itocon," in *DPSWS2020*, 2020.

[9] R. Takahashi, K. Hayashi, Y. Mitsukude *et al.*, "Itocon - a system for visualizing the congestion of bus stops around ito campus in real-time," in *The 18th ACM Conference on Embedded Networked Sensor Systems (SenSys 2020)*, 2020.

[10] S. Inoue, R. Takahashi, K. Hayashi *et al.*, "itocon:bus stop congestion visualization system using multiple congestion sensors," in *SeMI*, 2020.

[11] M. A. Abdelwahab, S. Kaji, M. Hori *et al.*, "Measuring "nigiwai" from pedestrian movement," *IEEE Access*, vol. 9, pp. 24 859–24 871, 2021.

[12] M. Terada, T. Nagata, and M. Kobayashi, "Population estimation technology for mobile spatial statistics," *NTT DOCOMO Technical Journal*, vol. 14, pp. 10–15, 2013.

[13] H. Arai, N. Ito, and Y. Taniguchi, "Image processing techniques for capturing crowds at a macroscopic level - estimating the number of people based on geometric models of people and crowds and its applications-," *CVIM*, vol. 2014, no. 13, pp. 1–8, 2014.

[14] T. Matsumoto, R. Takahashi, S. Ishida *et al.*, "Initial evaluation of device-free indoor congestion estimation using wireless LAN," *The 82nd National Convention Lecture and Paper Series*, vol. 2020, no. 1, pp. 247–248, 2020.

[15] E. Yuyama and H. Morino, "A study on a train congestion estimation system using infrared sensors and carbon dioxide sensors," *Tokyo Section Student Association, Institute of Electronics, Information and Communication Engineers*, vol. 53, 2009.

[16] C. Matte, M. Cunche, F. Rousseau *et al.*, "Defeating mac address randomization through timing attacks," in *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks (ACM WiSec)*, July 2016, pp. 15–20.