Optimizing intra-facility crowding in Wi-Fi environments using continuous-time Markov chains

Shinya Mizuno (s.mzn.eng@gmail.com)

Department of Computer Science, Shizuoka Institute of Science and Technology, 2200-2, Toyosawa, Fukuroi, Shizuoka, 437-8555

Haruka Ohba

Department of Computer Science, Shizuoka Institute of Science and Technology, 2200-2, Toyosawa, Fukuroi, Shizuoka, 437-8555

Research Article

Keywords: Optimization, Markov chain, Wi-Fi log, Parallel computing, K-means, Clustering, Facility operators, Crowdedness

Posted Date: June 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1441836/v3

Optimizing intra-facility crowding in Wi-Fi environments using continuous-time Markov chains

Shinya Mizuno*, Haruka Ohba

Department of Computer Science, Shizuoka Institute of Science and Technology, 2200-2, Toyosawa, Fukuroi, Shizuoka, 437-8555, Japan

*Corresponding author: Shinya Mizuno
Shizuoka Institute of Science and Technology
2200-2, Toyosawa, Fukuroi, Shizuoka, 437-8555, Japan
Email: s.mzn.eng@gmail.com

Abstract

Recently, increased measures are being devised to reduce crowdedness as a countermeasure for the spread of COVID-19. In this study, we propose a solution to reduce intra-facility crowdedness based on the usage of Wifi networks. This study maximizes the Wi-Fi logs that are continually generated in vast quantities in the ever-expanding Wi-Fi network environment to calculate the transition probabilities between nodes and the mean stay time at each node. Then, we model this data as a continuous-time Markov chain to obtain the variance of the stationary distribution, which we use as a metric of intra-facility crowdedness. Therefore, to minimize intra-facility crowdedness, we solved the optimization problem using stay rate as a parameter and demonstrated a numerical solution. In the optimization results, we succeeded in reducing

intra-facility crowding by approximately 30%. This solution is a realistic approach for reducing intra-facility crowdedness as it makes adjustments to people's stay times without any changes in their movements. We used the k-means method to categorize Wi-Fi users into a set of classes and documented the behavioral characteristics of each class, which helped implement class-specific measures to reduce intra-facility crowdedness. This enables facility managers to implement fine-grained countermeasures against crowdedness according to circumstances. Herein, we describe our computing environment and workflow for the basic analysis of vast quantities of Wi-Fi logs. Because the data we used are general-purpose, we believe that this research will be useful for both analysts and facility operators.

Keywords: Optimization, Markov chain, Wi-Fi log, Parallel computing, K-means, Clustering, Facility operators, Crowdedness

1. Introduction

The COVID-19 pandemic has forced governments, municipalities, and other institutions worldwide to implement countermeasures against the transmission of the disease. In addition to hygiene countermeasures, such as mouth rinsing and hand washing, it is important to stop crowd gatherings [1]. Although the increasing use of online platforms caters to this need, it has had a subsequent impact on making changes to existing physical environments. This facilitates the need for objective indicators to assess the crowdedness of a given facility.

Measuring crowdedness is important not only for COVID-19 countermeasures, but also for the optimization of crowdedness, which helps eliminate imbalances in the number of customers congregating in a certain region of interest. In the transportation field, the effects of crowdedness

have been documented from the perspectives of accessibility and mobility [2]. In recent years, there has been a growing interest in the use of cameras to study the movements of people in buses and other similar settings [3, 4]. However, because cameras capture only a specific point in space, it is difficult to analyze simultaneous events at multiple points.

Global Internet usage reached approximately 63% in 2021, making it an indispensable part of our daily lives [5]. The increasing smartphone penetration globally is mainly driven by increased internet connectivity over mobile networks as well as Wi-Fi. The diverse applications of the Wi-Fi environment, including its use in education, tourism, and disaster management and prevention, have led to the establishment of national Wi-Fi policies in several countries across the world. With the advent of 5G and other technologies, the Wi-Fi environment will be the most preferred network environment and will be used widely in virtual reality, remote offices, online classes, and many other fields, such as smart homes, support for the elderly, and nursing care [6–9].

In this study, we propose an optimization method to avoid crowding in facilities as a countermeasure against COVID-19. Our method analyzes the Wi-Fi logs that are created on a regular basis across Wi-Fi environments to study the behavior of facility users. In the proposed method, the constraints are set in a manner that allows users and facility managers to take realistic steps against crowdedness without drastically changing the status quo. First, we performed a basic analysis of the Wi-Fi users from the Wi-Fi log statistics. We classified users by factors such as frequency of use, location, and stay time, and identified the user groups that were present in the facility. Then, for each user group, we obtained information on the transition probabilities between nodes and the stay time at each node, and we modeled them as a continuous-time Markov chain. If a stationary distribution was obtained, the variance of the stationary distribution was used as a metric for assessing intra-facility crowdedness. We then

performed optimization under certain constraints, with the stay time parameter of each node as the explanatory variable and the variance of the stationary distribution as the objective function. This approach can help alleviate crowdedness under realistic conditions. By making effective use of WiFi environment logs, this study presents an effective optimization method for facility operators who are struggling to deal with COVID-19.

1.1 Related work

There has been significant research on the acquisition of user behavior using Wi-Fi logs in various fields [10–20]. For tourism behavior analysis, in particular, various studies using Wi-Fi logs have been conducted to understand the state of tourism behavior through facility and visitor stay times and OD (Origin–Destination) Tables [21, 22]. For example, the Internet-of-Things (IoT) data sent over Bluetooth and Wi-Fi have been used to monitor and estimate the number of passengers and the waiting time for buses and subway trains [23, 24]. Wi-Fi data have also been used to track tourists and score the attractiveness of tourist attractions [25]. This allows a strategic implementation of services, including the estimation of factors that impact tourism and offering recommendations to tourists [26, 27]. However, to date, this has not accelerated the use of Wi-Fi data for strategic planning.

In previous research, Markov chains were used to understand tourism behavior [28, 29]. These studies were based on the concept of absorbing Markov chains, wherein the user arrives from outside, and did not consider properties such as reachability or stationary distribution. In addition, because the volume of Wi-Fi log data is enormous, data-cleaning methods have been examined [30]. However, the computing environment and methods for analyzing vast quantities of log data have not been discussed extensively. The computing environment for handling ever-

increasing log data is significant for achieving real-time results. The Wi-Fi logs provided by vendor tools typically include basic information, such as the number of user connections, connecting devices, authentication status, and usage status. Although it is possible to obtain the arrival rate per unit time and the stay time of each user, the vendor-supplied tools do not track user transitions through access points, which limits their ability to offer a detailed picture of user trends.

To mitigate the impact of the COVID-19 pandemic, several efforts have been taken to investigate and predict the infection and movement of people [31]. Ainslie et al. [32, 33] found a very strong correlation between intra-urban migration and infection rate using intra-city movement data during the early stages of the pandemic. The use of susceptable, infected, and recovered (SIR) models to model and predict infections has been promoted [9, 34–36].

The use of IoT devices to monitor and intervene in public health in densely populated areas has also been investigated. A recent study incorporated the machine learning approach to maximize the testing resources to track people who have come into contact with an infected person [37]. Wang et al. [38] found that the interventions that focused on highly-mobile individuals and popular locations rather than the movements of actual people captured by Wi-Fi and GPS could reduce both peak infection rates and the total number of infected people, while maintaining high social activity levels. However, none of these studies have attempted to optimize crowdedness.

In this study, we present an environment wherein Wi-Fi logs can be analyzed realistically, and propose a method for defining and optimizing intra-facility crowdedness. Our approach enables the implementation of realistic anti-crowding measures and addresses the issues surrounding the global COVID-19 pandemic.

2. Materials and Methods

In this study, we used Wi-Fi logs to optimize intra-facility crowdedness. This section explains our approach for processing vast quantities of Wi-Fi logs and the computing environment needed. Then, using the characteristic quantities obtained from the processed Wi-Fi logs, we calculated the characteristics of different users and organized them into groups. Further, we used these user groups to assess intra-facility crowdedness using a continuous-time-type Markov chain. Finally, we propose an optimization model that minimizes the intra-facility crowdedness. The optimization model uses the stay time of the user as a parameter and imposes a fixed limit on the variability of the stay time, making it a realistic solution.

2.1. Process from Wi-Fi log to transition probability matrix calculation

2.1.1. Items used in the Wi-Fi logs

For a basic authentication at a Wi-Fi access point, the WPA2-Enterprise with 802.1X authentication and WPA2-PSK (shared network key) were used as the main security nodes. These authentication methods require the following log items, which are collected by most logging programs and used in this analysis: connection time, destination access point, and unique user identifier.

2.1.2. Overview of the Wi-Fi log data used

For our Wi-Fi log data, we used the Crawdad.org dataset, which is a dataset of association records for the Eduroam network at the KTH campuses, collected during 2014–2015 [39]. Table

1 lists the access point (AP) information for this dataset. This includes AP name and location information. In total, 1,123 APs were listed in the database.

Table 2 lists the content of the file containing user connection information. This file contains a unique user identifier, the AP to which the user is connected, and the connection time.

The files in the dataset were organized on a monthly basis, as shown in Table 3. The data containing N/A and users with only one connection were excluded. We targeted users who had connected more than once and had transitions between APs.

2.1.3. Calculation of transition probabilities and stay times

The process for calculating the transition probabilities and stay times from the Wi-Fi logs is as follows. This workflow is general-purpose for WiFi logs.

A. Preprocessing for Wi-Fi log analysis

(1) Delete records containing “N/A” from the original file.

(2) Delete users who appear only once in the original file.

(3) Output the files processed in (1) and (2).

B. Main processing for Wi-Fi log analysis

(1) Based on the file created in A (preprocessing), calculate the transition probabilities by parallel computation (MPI).

(2) Each MPI process calculates the number of transitions and stay times of the specified user and then outputs the results to a file with the aggregated information of all users.

C. Post-processing for Wi-Fi log analysis

(1) From the file obtained in B (main processing), delete APs that have no transitions, and calculate the transition probability matrix.

(2) Calculate the transition probability matrix and mean stay time for each classroom, floor, and building, with the APs aggregated, and output the results to a file.

Fig. 1 illustrates the state transition obtained in Section B. (Main process for calculating transition probabilities).

2.1.4. Process for calculating user groups

We performed clustering using the k-means method to study the characteristics of Wi-Fi users. From the results of the calculation of transition probabilities and stay times, we created a dataset for each user in the format shown below. Here, the node number indicates the number assigned when the AP aggregation is performed. Table 4 shows the environment in which clustering was computed.

User i = {node 1: number of connections …, node N: number of connections, node 1: stay time …, node N: stay time}

2.2. Computing environment for Wi-Fi log analysis

The Wi-Fi logs are enormous, and because the number of access points and users increases, the computation time increases. Therefore, a simple computing environment is inadequate for a seamless operation of businesses. In this study, we used a parallel computing environment to improve the efficiency of the Wi-Fi log computations. The programming environment used in

this study is listed in Table 5. The SQUID computing environment [40] at the Cybermedia Center of Osaka University was used. The file for 2014/09, which possessed the largest number of records, required a computation time of 8836.68 seconds. Depending on the availability of computing resources, the computation time can be further reduced by increasing the amount of parallelism.

2.3. Measuring intra-facility crowdedness from Wi-Fi logs using continuous-time Markov chains

In this section, we first define facility crowdedness using a continuous-time Markov chain [41]. We assume a finite state space S and continuous-time stochastic process $\{X(t); t \geq 0\}$. We define a continuous-time Markov chain with transition probability $\boldsymbol{P}(t) = \left(p_{ij}(t)\right), i, j \in S$ on S. Here, $X(t)$ is assumed to satisfy the following equation and be synchronous:

$$p_{ij}(t) = P(X(s+t) = j | X(s) = i), i, j \in S$$

In addition, for each transition probability $p_{ij}(t)$,

$$q_i = \lim_{h \downarrow 0} \frac{1 - p_{ii}(h)}{h} \in [0, \infty], i \in S$$

$$q_{ij} = \lim_{h \downarrow 0} \frac{p_{ij}(h)}{h} \in [0, \infty], i \in S, i \neq j$$

By assuming $q_{ii} = -q_i, i \in S$, we define the transition rate matrix below, where $\boldsymbol{P}(0) = \boldsymbol{I}$.

$$\boldsymbol{Q} = (q_{ij}) = \lim_{h \downarrow 0} \frac{\{\boldsymbol{P}(h) - \boldsymbol{I}\}}{h}$$

At time t when $X(t) = i$ and $i \in S$, the probability that the remaining stay time $\tau_i(t)$ is greater than u is given by the following equation:

$$P(\tau_i(t) > u | X(t-) \neq i, X(t) = i) = e^{-a_i u}$$

where $\frac{1}{a_i}$ is the mean stay time for $i \in S$. Therefore, the transition rate q_{ij} can be expressed as

$$\boldsymbol{q_{ij}} = \begin{cases} -\boldsymbol{a_i}\ (\boldsymbol{i} = \boldsymbol{j}) \\ \boldsymbol{a_i p_{ij}}\ (\boldsymbol{i} \neq \boldsymbol{j}) \end{cases} \tag{1}$$

When $X(t)$ is irreducible and ergodic, there is a limit distribution for $j \in S$

$$\pi_j = \lim_{t \to \infty} p_{ij}(t) \geq 0, \sum_{j \in S} \pi_j = 1$$

where π_j satisfies

$$\sum_{\boldsymbol{i} \in \boldsymbol{S}} \boldsymbol{\pi_i q_{ij}} = \boldsymbol{0}, (\boldsymbol{j} \in \boldsymbol{S}) \tag{2}$$

and $\{X(t)\}$ is the stationary distribution.

To define intra-facility crowdedness, we let σ be the variance of the stationary distribution of each state. Thus we obtain,

$$\sigma(\boldsymbol{a}) = \frac{1}{|S|} \sum_{i \in S} (\pi_i - \bar{\pi})^2$$

where $\bar{\pi} = \frac{1}{|S|} \sum_{i \in S} \pi_i$.

2.4. Method for optimizing intra-facility crowdedness

Next, we propose a method to optimize the crowdedness of the facilities [42]. In this section, we classify users according to their usage, and we introduce a set C of user classes. Because intra-facility crowdedness is defined as the variance of the stationary distribution of each state, we can reduce the intra-facility crowdedness by minimizing this variance, as shown in Equation (3). Here, the sum of the mean stay times in the facility is set as a constant value for each user class, as shown in Equation (4). This is done to avoid any hindrances in the use of the facility. Here, $\boldsymbol{a}'^{(\boldsymbol{c})}$ is the initial value of $\boldsymbol{a}^{(\boldsymbol{c})}$, $\pi_i^{(c)}\left(a_i^{(c)}\right)$ is the stationary distribution, given the stay rate $a_i^{(c)}$,

node i, and user class c, and $\overline{\pi}\left(a_i^{(c)}\right)$ is the mean value of the stationary distribution. Because a drastic change in the stay rate of users may cause confusion, we limit the change to a certain range for each user class, as shown in Equation (5).

$$\boldsymbol{Minimize}\ \boldsymbol{\sigma}\left(\boldsymbol{a}^{(1)}, \dots, \boldsymbol{a}^{(c)}, \dots\right) = \frac{\mathbf{1}}{|\boldsymbol{S}|}\sum_{c\in C}\sum_{i\in S}\left(\boldsymbol{\pi}_i^{(c)}\left(\boldsymbol{a}_i^{(c)}\right) - \overline{\boldsymbol{\pi}}\left(\boldsymbol{a}_i^{(c)}\right)\right)^2 \quad (3)$$

$$Subject\ to\ a_i^{(c)} \geq 0, i \in S, c \in C$$

$$\sum_{i\in S}\frac{1}{a_i^{(c)}} = K^{(c)}, \qquad K^{(c)} \in \mathbb{R} \quad (4)$$

$$\boldsymbol{a}'^{(c)}\left(1-\gamma^{(c)}\right) \leq \boldsymbol{a}^{(c)} \leq \boldsymbol{a}'^{(c)}\left(1+\gamma^{(c)}\right), \gamma^{(c)} \in \mathbb{R} \quad (5)$$

Facility crowdedness optimization algorithm

The facility crowdedness optimization algorithm is as follows.

1) Classify users into classes C based on facility Wi-Fi logs
2) For each class $c \in C$, give the transition probability matrix $\boldsymbol{P}^c(t) = \left(p_{ij}^c(t)\right), i, j \in S, c \in C$ and set the initial value of the stay time parameter $a_i^{(c)}$
3) Based on Equation (1), create the transition rate matrix $\boldsymbol{Q}^{(c)}$, and from Equation (2), obtain $\pi_i^{(c)}$
4) Solve the optimization problem (3) and find an $\boldsymbol{a}^{(c)}$ that satisfies the conditions
5) Repeat steps 3) and 4) to calculate the optimal $\boldsymbol{a}^{(c)}$

Table 6 shows our computing environment for optimizing intra-facility crowdedness.

3. Results and Discussion

3.1. Basic user analysis obtained from Wi-Fi logs

We categorized users into five classes based on the number of connections, location, and stay time. The characteristics of each class are listed in Table 7. Fig. 2 shows the number of users who used Wi-Fi in each building at least once. Most users belonged to Class 0 and had a short mean stay time. Because the data used in this study were obtained from Eduroam, which is a Wi-Fi roaming service that allows access from outside the university where it is installed, these could easily be outside users or students who do not use the network much.

For an accurate classification of the characteristics of the user classes, we clustered the buildings using the mean stay times of the users in each building (mean stay time for user class 0, mean stay time for user class 1, ..., mean stay time for user class 4). The results are presented in Table 8. Except for user class 0, more than 75% of all user classes were present in building cluster 2 at least once. The table reveals the characteristics of each user class—user class 2 uses building cluster 4, user class 3 uses building cluster 2, and so on.

3.2. Evaluation of intra-facility crowdedness by class

We evaluated the intra-facility crowding in each class. Table 9 shows the class-specific values related to the intra-facility crowding. The mean stay time (h) was the mean value of the stay time in each building for each class. The total mean stay time (h), which is the sum of the stay times in each building, is a constraint that is imposed on the optimization. The intra-facility crowdedness was computed for each class. Class 3 has the largest mean stay time, the largest sum thereof, and the greatest intra-facility crowdedness.

Figs. 3 and 4 show the stay rate and stationary distribution for each class, respectively.

3.3. Implementation of intra-facility crowdedness optimization

The preoptimized intra-facility crowding for all users was 0.00157494. The optimization computing environment is shown in Table 5. The number of iterations was 1502, and the computation time was 3196.20 seconds. After optimization, the intra-facility crowding was 0.00109697. Because the total mean stay time is a constraint that does not change, we checked the mean and standard deviation of the stay rate and the change in the intra-facility crowdedness in each class, as shown in Table 10. The table shows that optimization resulted in a decrease in the overall intra-facility crowdedness as well as the cluster-specific intra-facility crowdedness.

Fig. 5 shows the change in the stay rate before and after optimization. The negative values indicate buildings wherein the stay rate increased. The stay rates for Building 1 decreased compared to the values in Figs. 3 and 4, which means that for all classes, the stay time is longer and the value of the stationary distribution is greater. To standardize the value of the stationary distribution of this building, the optimization increases the stay rate and shortens the stay time, as shown in Fig. 5. Furthermore, it reduces the stationary distribution, as shown in Fig. 6. In Building 30, the stay rate was high, and the stationary distribution was low. After optimization, it was possible to increase the stationary distribution by decreasing the stay rate in classes 0, 2, and 4, and increasing it in classes 1 and 3. Thus, even for the same building, the increase or decrease in the stay rate differs according to the class, demonstrating that this method can provide realistic optimization results.

3.4 Features of this study

This study elaborated on the environment and process of analyzing huge amounts of Wi-Fi logs and calculating their statistics. A comparison with the previous study is shown in Table 11. The left side of the table shows the items related to the scale of the Wi-Fi logs, and the right side shows the items related to the analysis. In addition to basic statistical methods, clustering of users is also important for analyzing the logs to understand the trend of Wi-Fi users. Further optimization can provide an improved environment. In contrast to previous studies, this study presents a sufficient amount of Wi-Fi logs, as well as a log calculation flow and environment. As for the analysis method, we believe that it is academically and practically significant that we are able to optimize the system by using the classification of users as well as statistical analysis.

4. Conclusion

In this study, we presented a computational algorithm and its environment for effective use of huge Wi-Fi logs and classifying Wi-Fi users on the basis of clustering. We also proposed an optimization model by applying the transition probability matrix and stay rate obtained from Wi-Fi logs to a continuous-time Markov chain. This optimization model is effective in preventing intra-facility crowding, which was verified by numerical calculations. The model was able to reduce the crowding in the facility without changing the transition probability matrix, that is, without changing the flow line of people and only changing the stay rate. In addition, as the optimization is possible for each user class, the model is easy to adopt for facility management. By using this optimization model, we expect to utilize Wi-Fi logs to prevent user crowding and simultaneously increase the effectiveness of facility operations in COVID-19 prevention.

The limitation of this study is that there is no disconnection time available in the Wi-Fi logs. Therefore, we have set 3 hours as the maximum time spent. If there is no Wi-Fi cutoff time,

the accuracy can be improved by performing survival time analysis [43] on the time spent. The objective function of the optimization model is the variance of the stationary distribution; Osaki [42] uses an objective function that takes into account the facility area and the number of people the facility can accommodate, and it will be necessary to compare this with an objective function that includes the structure of the facility.

List of abbreviations

AP: access point

GPS: global positioning system

IoT: Internet of Things

MPI: parallel computation

OD: origin–destination

SIR: Susceptable, Infected, and Recovered model

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the CRAWDAD project, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of CRAWDAD.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by JSPS KAKENHI (grant number JP21K11774).

Authors' contributions

Mizuno conducted an analysis of Wi-Fi logs to build an optimization model. The main part of the paper is written by Mizuno.

Ohba was mainly in charge of the analysis of Wi-Fi logs and classification of users.

All authors read and approved the final manuscript.

Acknowledgments

Ljubica Pajevic, Gunnar Karlsson, Viktoria Fodor, CRAWDAD dataset kth/campus (v. 2019-07-01), traceset: eduroam, downloaded from https://crawdad.org/kth/campus/20190701/eduroam, https://doi.org/10.15783/c7-5r6x-4b46 5r6×4b46, July

This work was partly achieved through the use of SQUIDs at the Cybermedia Center, Osaka University.

References

1. Hatabu A, Mao X, Zhou Y, Kawashita N, Wen Z, Ueda M, et al. Knowledge, attitudes, and practices toward COVID-19 among university students in Japan and associated factors: An online cross-sectional survey. PLOS ONE. 2020;15:e0244350. https://doi.org/10.1371/journal.pone.0244350
2. Stopher PR. Reducing road congestion: A reality check. Transp Policy. 2004;11:117–31. https://doi.org/10.1016/j.tranpol.2003.09.002
3. Yahiaoui T, Meurie C, Khoudour L, Cabestaing F. A people counting system based on dense and close stereovision. In: Elmoataz A, Lezoray O, Nouboud F, Mammass D, editors. Image and signal processing. Berlin, Heidelberg: Springer; 2008. p. 59–66
4. Eshel R, Moses Y. Tracking in a dense crowd using multiple cameras. Int J Comput Vis. 2010;88:129–43. https://doi.org/10.1007/s11263-009-0307-0
5. Glueck DH, Karimpour-Fard A, Mandel J, Muller KE. Probabilities for separating sets of order Statistics. https://www.itu.int:443/en/ITU-D/Statistics/Pages/stat/default.aspx. Accessed 2021-12-29. Statistics (Ber). 2010;44:145–53. https://doi.org/10.1080/02331880902986984
6. Information and communications in Japan 2018. https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2019/2019-index.html. Accessed 2021-12-29

7. Abdul Kader M, Nayim Uddin M, Mohammad Arfi A, Islam N. Anisuzzaman Md. Design amp; Implementation of an Automated Reminder Medicine Box for Old People and Hospital. Eng Technol (ICISET). 2018 International Conference on Innovations in Science; 2018. p. 390–4
8. Li B, Yu J. Research and application on the smart home based on component technologies and Internet of things. Procedia Eng. 2011;15:2087–92. https://doi.org/10.1016/j.proeng.2011.08.390
9. Yang M, Huang H, Yuan H, Sun Q. Interaction design of products for the elderly in smart home under the mode of medical care and pension. In: Zhou J, Salvendy G, editors. Human aspects of IT for the aged population healthy and active aging. Cham: Springer International Publishing; 2016. p. 145–56
10. Hwang I, Jang YJ. Process mining to discover shoppers' pathways at a fashion retail store using a WiFi-base indoor positioning system. IEEE Trans Autom Sci Eng. 2017;14:1786–92. https://doi.org/10.1109/TASE.2017.2692961
11. Kianoush S, Savazzi S, Rampa V, Nicoli M. People counting by dense WiFi MIMO networks: Channel features and machine learning algorithms. Sensors (Basel). 2019;19:3450. https://doi.org/10.3390/s19163450
12. Trivedi A, Zakaria C, Balan R, Becker A, Corey G, Shenoy P. WiFiTrace: Network-based contact tracing for infectious diseases using passive WiFi sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol. Proceedings of the ACM InterAct mobile wearable ubiquitous technique. 2021;5:1–26. https://doi.org/10.1145/3448084

13. Meneses F, Moreira A. Large scale movement analysis from WiFi based location data; 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN). 2012. p. 1–9

14. Vu L, Nguyen P, Nahrstedt K, Richerzhagen B. Characterizing and modeling people movement from mobile phone sensing traces. Pervasive Mob Comput. 2015;17:220–35. https://doi.org/10.1016/j.pmcj.2014.12.001

15. Zola E, Barcelo-Arroyo F. A comparative analysis of the user behavior in academic WiFi networks. Proceedings of the 6th ACM workshop on performance monitoring and measurement of heterogeneous wireless and wired networks 2011;Oct 31:(59–66)

16. Redondi AE, Cesana M, Weibel DM, Fitzgerald E. Understanding the WiFi usage of university students; 2016. International wireless communications and mobile computing conference (IWCMC) [Internet] Cited 2022 Feb 19. http://ieeexplore.ieee.org/document/7577031/. Paphos, Cyprus: IEEE Publications. p. 44–9

17. Basalamah A. Crowd mobility analysis using WiFi sniffers. IJACSA [Internet]; 2016 Cited 2022 Feb 19. http://thesai.org/Publications/ViewPaper?Volume=7&Issue=12&Code=ijacsa&SerialNo=49. 7

18. Danalet A, Tinguely L, de Lapparent Md, Bierlaire M. Location choice with longitudinal WiFi data. J Choice Modell. 2016;18:1–17. https://doi.org/10.1016/j.jocm.2016.04.003

19. Ruiz-Ruiz AJ, Blunck H, Prentow TS, Stisen A, Kjærgaard MB. Analysis methods for extracting knowledge from large-scale WiFi monitoring to inform building facility planning. IEEE Int Conf Pervasive Comput Commun (PerCom). 2014:(130–8)

20. Nakanishi W, Kobayashi H, Tsuru T, Matsumoto T, Tanaka K, Suga Y, et al. Understanding travel pattern of tourists from Wi-Fi probe requests: A case study in Motobu Peninsula, Okinawa. JSCE Ser D3. 2018;74:I_787–97. https://doi.org/10.2208/jscejipm.74.I_787

21. Ryu S, Park BB, El-Tawab S. WiFi sensing system for monitoring public transportation ridership: A case study. KSCE J Civ Eng. 2020;24:3092–104. https://doi.org/10.1007/s12205-020-0316-7

22. Nguemdjo U, Meno F, Dongfack A, Ventelou B. Simulating the progression of the COVID-19 disease in Cameroon using SIR models. PLOS ONE. 2020;15:e0237832. https://doi.org/10.1371/journal.pone.0237832

23. Shlayan N, Kurkcu A, Ozbay K. Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals. 2016 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE Publications; 2016. p. 229–34

24. El-Tawab S, Yorio Z, Salman A, Oram R, Park BB. Origin-destination tracking analysis of an intelligent transit bus system using Internet of things. 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); 2019. p. 139–44

25. Arreeras T, Arimura M, Asada T, Arreeras S. Association rule mining tourist-attractive destinations for the sustainable development of a large tourism area in Hokkaido using Wi-Fi tracking data. Sustainability. Multidisciplinary digital publishing institute. 2019;11:3967

26. Gavalas D, Kasapakis V, Konstantopoulos C, Mastakas K, Pantziou G. A survey on mobile tourism Recommender Systems. 2013 Third International Conference on Communications and Information Technology (ICCIT); 2013. p. 131–5

27. Arreeras T, Endo M, Takahashi H, Asada T, Arimura M. An association rule mining-based exploration of travel patterns in wide tourism areas using A Wi-Fi package sensing survey. J East Asia Soc Transp Stud. 2019;13:1099–113

28. Gabe A, Kamiya D, Yamanaka R, Fukuda D, Suga Y. Trial of Wi-Fi packet sensing for surveying tourists travel behavior remote islands: A Case Study in Yaeyama Islands. Proceedings of the fuzzy system symposium. 35. Japan Society for Fuzzy Theory and Intelligent Informatics; 2019. p. 479–81. (in Japanese)

29. Dantsuji T, Sugishita K, Fukuda D, Asano M. Analysis of the properties of tourists' dwell time using Wi-Fi packet data:A case study of the approach to Hase-dera temple. J City Plan Inst Jpn. 2017;52:247–54. (in Japanese). https://doi.org/10.11361/journalcpij.52.247

30. Alessandrini A, Gioia C, Sermi F, Sofos I, Tarchi D, Vespe M. WiFi positioning and Big Data to monitor flows of people on a wide scale. 2017 Eur Navig Conference (ENC); 2017. p. 322–8

31. COVID-19 impacts on transport. https://opendata-nzta.opendata.arcgis.com/datasets/covid-19-impacts-on-transport, Accessed 2021-12-21

32. Ainslie KEC, Walters CE, Fu H, Bhatia S, Wang H, Xi X, et al. Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. Wellcome Open Res. 2020;5:81. https://doi.org/10.12688/wellcomeopenres.15843.2

33. Conrow L, Campbell M, Kingham S. Transport changes and COVID-19: From present impacts to future possibilities. N Z Geogr Cited 2021 Dec 21. 2021;77:185–90. https://onlinelibrary.wiley.com/doi/abs/10.1111/nzg.12315. https://doi.org/10.1111/nzg.12315

34. Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D. Early prediction of the 2019 novel coronavirus outbreak in the Mainland China based on simple mathematical model. IEEE Access. 2020;8:51761–9. https://doi.org/10.1109/ACCESS.2020.2979599

35. Ndiaye BM, Tendeng L, Seck D. Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting; 2020 Accessed 2021 May 25. https://arxiv.org/abs/2004.01574v1

36. Simha A, Prasad RV, Narayana S. A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe; 2020. arXiv:200311920 [math, q-bio] [Internet] Accessed 2021 Dec 21. http://arxiv.org/abs/2003.11920

37. Ness B. Shroff, Zhiqiang Lin, Dong Xuan, Fang Liu, Wenbo Ren, Rahul Singh. A blueprint for effective pandemic mitigation. ITU J-FET. 2020;1:89–101. https://doi.org/10.52953/MJUY7335

38. Wang H, Ghosh A, Ding J, Sarkar R, Gao J. Heterogeneous interventions reduce the spread of COVID-19. In: Simul Real Mob Data Arxiv. 200806549 [physics]. http://arxiv.org/abs/2008.06549. Accessed 2021-12-29

39. Pajevic L, Karlsson G, Fodor V. CRAWDAD Dataset Kth/Campus (v.2019-07-01). CRAWDAD Wirel Network Data Arch. 2019. https://doi.org/10.15783/c7-5r6x-4b46. Accessed 2021-12-29

40. Cybermedia Center, SQUID. http://www.hpc.cmc.osaka-u.ac.jp/squid/. Accessed 2021-12-21 (in Japanese)

41. Miyazawa M. Kindai kagaku sha Co., Ltd, Probability and Probability Process (Modern Mathematics Seminar; 1993

42. Ohsaki M. Facility Layout Optimization based on Markov Chain Model and Genetic Algorithms. J Archit Plan (Trans AIJ). 1998;63:251–8. https://doi.org/10.3130/aija.63.251_1

43. Kleinbaum DG, Klein M. Survival analysis: A self-learning text. 3rd ed. New York: Springer Science+Business Media, LLC, https://doi.org/10.2307/2532873, 2012

Table 1. Excerpt from APlocations.csv (1,123 items in total)

AP	x_coordinate (m)	y_coordinate (m)	Floor
Bldg1AP1	21534	32313	2
Bldg1AP2	21534	32313	2
Bldg1AP3	21534	32313	3

Table 2. Excerpt from traceset/2014_01.csv

Timestamp	Client	AP
2014-01-01 00:00:24	b7f22f…	Bldg11AP21
2014-01-01 00:00:30	336501…	Bldg44AP3
2014-01-01 00:00:35	4b912f…	Bldg48AP65

Table 3. Summary of usage by month

Year/month	Total number of records	Number of users	Amount of data (Mbyte)
2014/01	5634970	27052	402.5
2014/02	6803432	27164	485.9
2014/03	6671451	28351	476.4
2014/04	6178580	28178	441.3
2014/05	6466906	30525	461.8
2014/06	3284709	21776	234.8
2014/07	1426694	8495	102.1
2014/08	3605161	18787	257.5
2014/09	7655584	33837	546.2
2014/10	7401092	34374	528.2
2014/11	7114147	34107	507.6
2014/12	5366380	31738	383.0
2015/01	5813530	32057	414.8
2015/02	6806318	32428	485.5
2015/03	7235589	33750	516.2
2015/04	5818219	32417	415.2

Table 4. Computing environment for user clustering

Item	Description
Programming language	Python 3.7.12
Library used	scikit-learn==1.0.1

Technique	K-Means
Number of clusters	5

Table 5. Computing environment for Wi-Fi log analysis

Item	**Description**
Programming language	Python 3.6.13
Library used	mpi4py 3.1.3
Computing environment	SQUID (Osaka University)
Processor information	Intel(R) Xeon(R) Platinum 8368 Packages(sockets): 2 Cores: 76 Processors (CPUs): 152 Cores per package: 38 Threads per core: 2
Memory	248 GB/Node
Number of nodes used	1
Number of cores used	76 (all in parallel)

Table 6. Environment for optimizing intra-facility crowdedness

Item	**Description**
Programming language	Python 3.7.12
Library used	scipy==1.4.1
Optimization algorithm	Constrained trust region method

Explanatory variable	Stay rate $a^{(c)}$ at each node
Gradient function	Gradient estimation (2-point) applied
Hessian matrix	Hessian estimation (BFGS) applied
$\boldsymbol{\gamma^{(c)}}$	$\gamma^{(c)} = 0.25$ in all classes

Table 7. Characteristics of each class

User class	**Number of users**	**Mean stay time (h)**	**Number of building nodes**			
			1/3 of cluster stays	Mean stay time < 1 h	Mean stay time < 2 h, ≥ 1 h	Mean stay time < 3 h, ≥ 2 h
0	23221	0.003	1	25	14	4
1	879	0.442	10	18	5	20
2	6476	0.050	22	12	5	26
3	2225	0.092	3	22	9	12
4	1036	0.345	15	15	6	22

Table 8. Clustering of building nodes

Building cluster 0	**Building cluster 1**	**Building cluster 2**	**Building cluster 3**	**Building cluster 4**
6, 7, 12, 13, 17, 18,	0, 5, 8, 10, 14, 15, 20,	1, 4	3, 11, 16, 23, 24, 25,	2, 9, 22

	19, 21, 31, 33		28, 32, 37, 38, 39, 40, 41, 42				26, 27, 29, 30, 34, 35, 36			
User class	rate (%)	time (h)	rate (%)	time (h)	rate (%)	time (h)	rate (%)	time (h)	rate (%)	time (h)
0	14.9	0.023	4.2	0.011	37.3	0.054	6.7	0.011	19.4	0.031
1	35.9	0.147	2.7	0.008	75.8	0.291	12.7	0.038	72.7	0.660
2	61.7	0.161	6.1	0.016	86.3	0.179	33.5	0.065	78.6	0.275
3	14.3	0.051	1.4	0.006	81.6	0.691	4.1	0.013	29.3	0.105
4	50.3	0.233	3.3	0.008	84.8	0.266	20.4	0.044	63.7	0.240

Table 9. Class-specific values related to intra-facility crowdedness

Class (c)	**Mean stay time (hours) ($Mean(a^{(c)})$)**	**Total mean stay time (hours) ($K^{(c)}$)**	**Intra-facility crowdedness**
Class 0	0.5056	21.74	0.000813
Class 1	0.8170	35.13	0.001513
Class 2	0.6227	26.77	0.001223
Class 3	0.8276	35.58	0.002853
Class 4	0.6617	28.45	0.001469

Table 10. Changes in stay rate and intra-facility crowdedness after optimization

	Before optimization	**After optimization**	**Before optimization**	**After optimization**	
Class (*c*)	Mean and standard deviation of stay rate	Mean and standard deviation of stay rate	Intra-facility crowdedness	Intra-facility crowdedness	Reduction rate (%)
Class 0	3.1941, 2.7725	2.9431, 2.6706	0.000813	0.000560	68.8806
Class 1	2.1562, 1.7184	2.1876, 1.9235	0.001513	0.001072	70.8526
Class 2	3.2556, 3.1473	2.8221, 2.7170	0.001223	0.000863	70.5641
Class 3	2.1756, 2.7388	2.4671, 3.4307	0.002853	0.001924	67.4377
Class 4	3.1551, 3.3967	2.7481, 3.1901	0.001469	0.001063	72.3621

Table 11. Comparison between this study and previous studies

Reference number	**Number of APs**	**Number of users (devices)**	**Number of data**	**Log computing environment and Algorithm**	**Statistical analysis**	**Clustering**	**Optimization**
This study	1,123	33,837	7,655,584	mentioned	mentioned	mentioned	mentioned
[11]	NA	NA	NA	NA	mentioned	NA	NA

[13] US	Approx. 5,500	Approx. 38,000	NA	mentioned	mentioned	mentioned	NA
[13] Singapore	Approx. 13,000	Approx. 50,000					
[14]	550	5,989	1,437,504	NA	mentioned	NA	NA
[15]	NA	123	43,150	NA	mentioned	mentioned	NA
[16]	37	7,336	NA	NA	mentioned	NA	NA
[17]	28	27,538	300,681	NA	mentioned	mentioned	NA
[18]	8	69,467	4,517,687	NA	mentioned	NA	NA

Figure Captions

Fig. 1 State transition in Wi-Fi logs

Fig. 2. Number of users who used the Wi-Fi in each building node at least once

Fig. 3. Stay rate for each class

Fig. 4. Stationary distribution for each class

Fig. 5. Change in stay rate in each class before and after optimization

Fig. 6. Change in stationary distribution for each class before and after optimization

Figures

Figure 1

State transition in Wi-Fi logs

Figure 2

Number of users who used the Wi-Fi in each building node at least once

Figure 3

Stay rate for each class

Figure 4

Stationary distribution for each class

Figure 5

Change in stay rate in each class before and after optimization

Figure 6

Change in stationary distribution for each class before and after optimization