Origin–Destination Estimation for Route Buses Using BLE Advertising Packets: A Case Study on Two Real-World Routes

Taito Yoshimura*, Kotaro Hayashi*, Ismail Arai†, Yuki Matsuda*‡,
* *Okayama University*, Okayama, Japan,
Email: {taito.yoshimura, kotaro.hayashi}@cocolab.jp, yukimat@okayama-u.ac.jp.
† *Nara Institute of Science and Technology*, Nara, Japan,
Email: ismail@is.naist.jp.
‡ *RIKEN Center for Advanced Intelligence Project AIP*, Tokyo, Japan.

***Abstract*—In recent years, regional bus operations have faced critical challenges including declining ridership and driver shortages. To address these issues and develop efficient operational plans, collecting Origin-Destination (OD) data is essential. However, conventional OD data collection methods are constrained by labor costs and time limitations, while IC card-based approaches miss flat-rate users and cash-paying passengers. Additionally, the introduction of facial recognition cameras raises social acceptability concerns due to surveillance implications. To address these challenges, this study focuses on using Bluetooth Low Energy (BLE) advertising packets emitted by passengers' mobile devices. However, modern devices periodically and randomly change their BLE MAC addresses, making continuous tracking difficult throughout a trip. To overcome this, we propose a bus OD data estimation method that applies MAC address carry-over technique. This approach exploits the asynchronous nature of MAC address changes and the embedded information in packets to enable continuous tracking. To evaluate the effectiveness of the proposed method, we conducted field experiments on two different operational bus routes. The results demonstrate that the proposed method can estimate OD data with an average precision of 70% (max: 100%, min: 53%) and an average recall of 34% (max: 59%, min: 15%).**

***Index Terms*—Origin-Destination Estimation, Bluetooth Low Energy (BLE), MAC Address Randomization**

I. Introduction

According to a document from Japan's Ministry of Land, Infrastructure, Transport and Tourism titled "Current Status and Issues Surrounding Regional Public Transportation and Perspectives for Consideration" [1], problems facing regional bus services include declining passenger numbers and driver shortages. Given this situation, bus operators need to operate efficiently with limited labor and vehicle resources. To achieve efficient operation, it is essential to understand the actual usage patterns of buses. Origin-Destination (OD) data, which records each passenger's boarding bus stop (Origin) and alighting bus stop (Destination), is effective for grasping usage patterns. By utilizing OD data, it becomes possible to take measures such as increasing services on high-demand sections and reducing them on low-demand sections, as well as reducing bus stops [2]. Furthermore, for sections with many transferring passengers, convenience and profitability can be improved by adding direct routes that eliminate the need for transfers.

Traditionally, mainstream methods for acquiring OD data included distributing cards to passengers by surveyors, visual recording by drivers or onboard surveyors, and methods using IC cards [3]–[5]. However, data collection through card distribution or visual observation is limited to a few days due to labor cost constraints, resulting in data for only a limited period. When collecting data over a long period and across multiple routes, personnel costs become a significant issue. Moreover, while methods using IC cards can confirm OD data from usage history, they have the limitation of not including users with flat-rate passes or those paying in cash. Other methods using cameras also exist, but they face challenges such as installation location constraints and concerns about social acceptability, like creating a sense of being monitored, making their easy adoption difficult. Considering these issues, there is a demand for establishing a low-cost OD data acquisition method that respects privacy.

Therefore, this research aims to establish a method for estimating OD data on route buses by scanning BLE (Bluetooth Low Energy) advertising packets emitted from mobile devices such as smartphones. In this paper, we propose a method for estimating OD data for route buses using BLE advertising packets. As a result of conducting field experiments on two different bus routes operating in Okayama City, we obtained results indicating that the proposed method can estimate OD data with an average precision of 70% (maximum 100%, minimum 53%) and an average recall of 34% (maximum 59%, minimum 15%).

II. Related Work

A. Existing OD Estimation Approaches

One approach to estimating OD data involves the use of cameras. For instance, Zhang *et al.* [6] proposed a passenger reidentification method for estimating route-level bus passenger OD flow using video images at bus doors. Camera-based

This study was supported in part by JST PRESTO under Grant No. JPMJPR2039 and JST COI-NEXT under Grant No. JPMJPF2115. We also thank Chutetsu Bus Co., Ltd. for their invaluable cooperation.

approaches can achieve high performance, but they often faces significant barriers to adoption. For example, concerns over surveillance and social acceptability make them difficult to implement.

Another promising line of research leverages wireless packets emitted by mobile phones. Ryu *et al.* [7] estimated OD data using only Wi-Fi packets. Kawashima *et al.* [8] employed BLE advertising packets for the same purpose. Ziyuan *et al.* [9] combined Wi-Fi and Bluetooth packets to enhance OD estimation accuracy. In these studies, sensing devices were installed inside buses to capture Wi-Fi and Bluetooth packets emitted by smartphones. From these packets, MAC addresses were extracted. Non-passenger MAC addresses were filtered based on factors such as the number of detections and average RSSI values. The appearance and disappearance times of MAC addresses were then mapped to the departure and arrival times at specific bus stops to estimate OD data. However, modern smartphones employ randomized MAC addresses for privacy protection. These addresses change periodically, making the aforementioned techniques difficult to apply in practice with current devices.

B. Research on Tracking Devices with Randomized MAC Addresses

Several studies have addressed the challenge of tracking devices that use randomized MAC addresses. Kawashima *et al.* [8] utilized BLE advertising packets generated by the Exposure Notification system developed by Google and Apple to track such devices and estimate OD data. Exposure Notification is a technology developed to help prevent the spread of COVID-19, which anonymously tracks contact with other users via Bluetooth. In their method, Rolling Proximity Identifiers and RSSI values were used to track BLE devices of bus passengers and estimate their OD data. However, this method relies on users having installed applications like COCOA (the COVID-19 contact-tracing app) [10]. As the service was discontinued in November 2022, the method is no longer applicable. Akiyama *et al.* [11] proposed a regression model that identifies devices with randomized MAC addresses by analyzing the timing and RSSI trends of received advertising packets. While their method achieved high accuracy, it performs poorly when MAC address changes occur simultaneously. Moreover, their study was limited to laboratory conditions and has not been validated in real-world environments. Becker *et al.* [12] proposed an "address carry-over algorithm" that associated newly randomized MAC addresses by exploiting a vulnerability where identifiers embedded in BLE advertising payloads did not change in sync with the MAC address. However, this vulnerability has since been addressed, rendering the method ineffective for current devices.

C. Position of This Study

In summary, existing methods for OD data collection in route buses face several challenges. Manual surveys are labor-intensive and costly, while sensor-based methods such as camera tracking raise privacy concerns. IC card–based methods fail to account for flat-rate and cash-paying passengers. Packet-based methods that rely on device emissions often do not accommodate MAC address randomization. While methods using COCOA-related packets do address this issue, they are no longer applicable due to the discontinuation of the service.

In this study, we aim to overcome these limitations by proposing a BLE-based OD estimation method that supports randomized MAC addresses.

III. Proposed Method

This section presents our proposed method for estimating OD data using BLE advertising packets, based on our preliminary experiments [13], [14]. An overview of the proposed method is shown in Fig. 1, and each step is described in detail below.

Step (1): Packet Collection
: We install BLE scanners at both the front and rear seats of the bus to capture BLE advertising packets. Each packet is recorded with the scanning timestamp, RSSI, randomized MAC address, and advertising (AD) data. Because the recorded MAC addresses are randomized, they are not directly linked to individuals, offering relatively low privacy risk. However, even randomized addresses may be re-identifiable when combined with other information. Therefore, to enhance privacy protection, the MAC addresses are transformed using a salted hash function, ensuring that the original values cannot be reverse-engineered.

Step (2): Filtering
: We extract device-specific packets using the following criteria: iPhones are identified by packets with Manufacturer Specific Data starting with `0x004c10`, and Android smartphones by Service Data UUIDs beginning with `0xfef3`. Next, we extract MAC addresses observed by both the front and rear scanners, assuming these devices belong to onboard passengers. To further eliminate packets from non-passenger devices (e.g., roadside pedestrians), we apply a filtering process based on the total number of detections from both scanners.

Step (3): Address Carry-Over
: Since MAC address changes occur asynchronously, we attempt to carry over identity between address changes using AD data and other packet features. Candidates for the next MAC address are selected based on a predefined threshold. If no candidates are found, the device is considered to have alighted or the data is treated as noise. If multiple candidates exist, the one with the closest average RSSI is selected as the most likely match.

Step (4): OD Data Estimation
: From the results of Step 3, we determine the appearance and disappearance times of each device, which are considered indicative of boarding and alighting, respectively. As shown in Fig. 2, if these

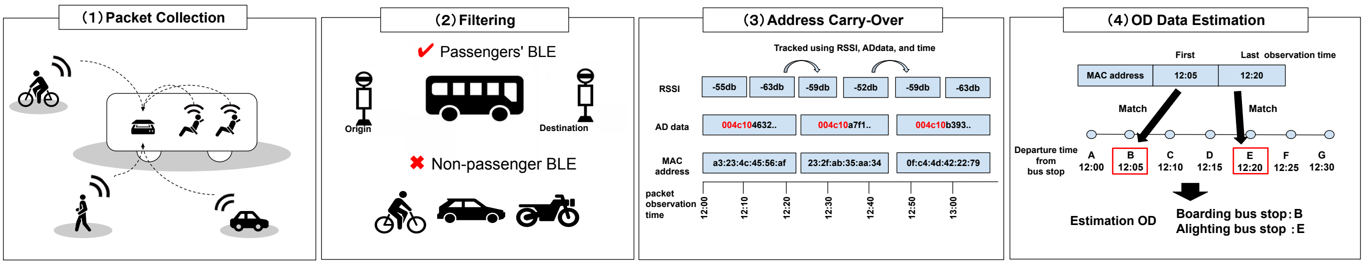

Fig. 1. Overview of the proposed method

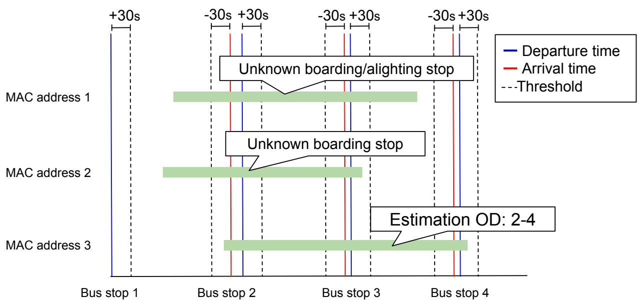

Fig. 2. Threshold for passenger judgment

timestamps fall within the threshold of bus departure or arrival times at a stop, the device is recognized as a passenger and OD data is generated. MAC addresses that do not meet this criterion, such as those with ambiguous boarding or alighting stops (e.g., MAC addresses 1 and 2 in Fig. 2), are excluded from the final OD dataset.

IV. Experiment and Evaluation

A. BLE Sensing Device

For BLE advertising packet collection (hereafter referred to as "scanner"), we used a Raspberry Pi 4 Model B equipped with a USB Bluetooth adapter (BUFFALO BSBT4D100, Bluetooth 4.0+EDR/LE Class 1 compatible) as shown in Fig. 3. Although the Raspberry Pi 4 has a built-in Bluetooth module, we opted for an external module to standardize the hardware specifications. To monitor scanner operation and maintain time synchronization, we integrated an LTE communication module—specifically, a USB LTE dongle (PIXELA PIX-MT110 or Fujisoft +F FS040U).

For BLE scanning, we employed the Python library bluepy[1]. BLE scanning refers to detecting nearby BLE devices. There are two types of scans: passive scan is a method that the scanner passively listens for and receives advertising packets, and active scan is a method that the scanner sends a scan request to the advertising device to obtain additional information after receiving an advertising packet.

[1]https://github.com/IanHarvey/bluepy

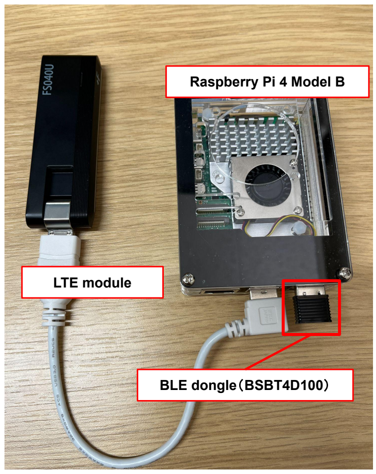

Fig. 3. BLE sensing devices

The bluepy library supports both modes[2], and in our experiments, we adopted the active scan mode. Given the challenges of securing stable power sources within a bus and the need for flexible installation, all scanners were powered using portable batteries.

B. Data Collection Routes

We selected two route buses operating within Okayama City for data collection. The routes are illustrated in Fig. 4, with the National Hospital Route shown in red and the Industrial Technology Center Route in blue.

These routes were chosen because each has a relatively long duration (over 30 minutes) and many bus stops. Since our OD estimation method relies on detecting changes in the randomized MAC addresses emitted by passengers' devices, longer ride durations increase the likelihood of observing such changes during a single trip. These routes are well-suited for verifying the effectiveness of the proposed method because their relatively long travel times increase the likelihood of MAC address changes during the ride. Additionally, the high number of stops allows for diverse Origin-Destination (OD) patterns, which supports more comprehensive evaluation.

To evaluate robustness across different environments, we selected two routes with distinct characteristics: The National

[2]While not explicitly documented, the scan function in bluepy allows switching between modes using the `passive=True` or `False` argument. The default is `passive=False`, i.e., active scan.

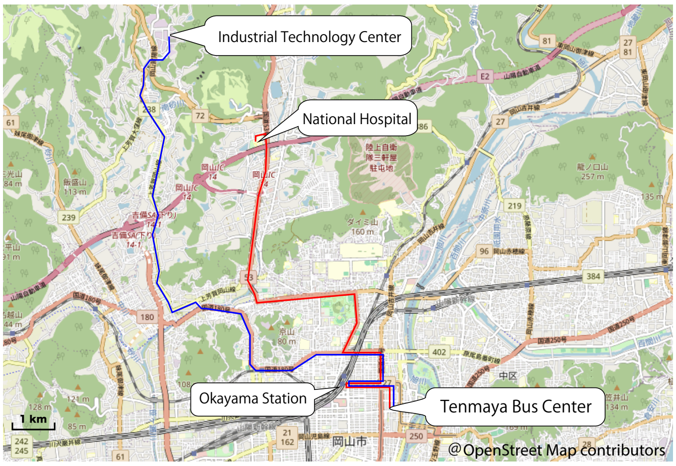

Fig. 4. Route map of the experimental area

Hospital Route passes through central Okayama City, where traffic is typically dense. In contrast, the Industrial Technology Center Route includes suburban residential areas and industrial parks, where traffic is generally lighter. This diversity in traffic conditions helps assess the method's generalizability. Based on these considerations, the two selected routes were deemed appropriate for validating the proposed method.

The experiments were conducted with the cooperation of Chutetsu Bus Co., Ltd., on October 15 and December 18, 2024. The following is a list of the actual datasets collected during the evaluation:

- National Hospital Route
  (Start: Tenmaya Bus Center / Terminus: National Hospital)
  - October 15, 8:25 - 9:08
  - October 15, 14:30 - 15:11
  - October 15, 17:37 - 18:31
- Industrial Technology Center Route
  (Start: Tenmaya Bus Center / Terminus: Industrial Technology Center)
  - December 18, 7:33 - 8:29
  - December 18, 11:10 - 11:51 (Terminus: Sayama Danchi)
  - December 18, 14:30 - 15:18
  - December 18, 17:20 - 18:11 (Terminus: Sayama Danchi)
- Industrial Technology Center Route
  (Start: Industrial Technology Center / Terminus: Tenmaya Bus Center)
  - December 18, 8:46 - 9:38
  - December 18, 11:58 - 12:36 (Start: Sayama Danchi)
  - December 18, 15:31 - 16:04 (Terminus: Okayama Station)

For each ride, BLE scanners were installed at the front (Scanner 1) and rear (Scanner 2) of the bus to collect data. The scan interval was set to 1 second. Each scan recorded the timestamp, randomized MAC address, RSSI, and AD data. Additionally, at every stop, the bus's stop and departure times were manually recorded—defined respectively as the time the doors opened and closed.

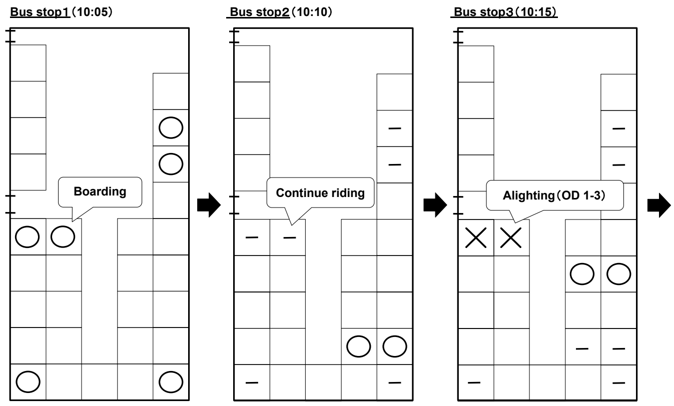

Fig. 5. Example of annotation memo

C. Ground Truth OD Data Collection

To create ground truth OD data for evaluation, we manually recorded the boarding and alighting stops of each passenger. Since we recorded data from general passengers, we could not determine whether their smartphones had Bluetooth turned on.

An example of a handwritten annotation is shown in Fig. 5. First, a seating chart corresponding to each bus stop was prepared. When a passenger boarded at a stop, a circle ($\circ$) was drawn at their seat; when a passenger alighted, a cross ($\times$) was drawn. For passengers who remained onboard, a hyphen or similar mark was used. In rare cases, especially during crowded times, passengers moved to different seats. These movements were indicated with arrows between seats. If standing passengers were observed due to congestion, their approximate locations and OD information were written in the margin of the seating chart. By tracking these seat-level annotations—from circles to crosses—accurate OD data was compiled, including both seated and standing passengers.

D. Evaluation Method

To evaluate the proposed method, we compared the estimated OD data with the ground truth OD data. For each bus operation, bus stops where the vehicle did not stop were excluded from the analysis.

The evaluation was based on three standard metrics: recall, precision, and F1-score. Recall represents the proportion of correctly estimated OD data among the ground truth data (Eq. 1). Precision represents the proportion of correct estimations among all OD data estimated by the proposed method (Eq. 2). A predicted OD is considered incorrect if either the boarding or alighting stop does not match the ground truth, or if there is no corresponding ground truth OD entry. In cases where the predicted number of OD matches exceeds the actual number (e.g., if three OD records are estimated for a route segment where only one person actually traveled), only the number matching the ground truth is counted as correct; the rest are considered false positives.

TABLE I
EVALUATION RESULTS FOR EACH OPERATION

	Time	# of Passengers	Precision	Recall
National Hospital Route	8:25-9:08	22	86%	59%
	14:30-15:11	19	75%	15%
	17:37-18:31	35	85%	31%
Industrial Technology Center Route	7:33-8:29	35	61%	31%
	8:46-9:38	23	53%	34%
	11:10-11:51	13	100%	23%
	11:58-12:36	17	54%	41%
	14:30-15:18	14	57%	28%
	15:31-16:04	8	66%	50%
	17:20-18:11	28	64%	32%

$$\text{Recall} = \frac{\text{Number of Correct OD Estimations}}{\text{Number of ground Truth ODs}} \quad (1)$$

$$\text{Precision} = \frac{\text{Number of Correct OD Estimations}}{\text{Number of Estimated ODs}} \quad (2)$$

E. Evaluation Results

Table I summarizes the performance of the proposed method for each bus operation. The method was evaluated using two different bus routes in Okayama City.

The results indicate that the proposed method achieved an average precision of 70% (with a maximum of 100% and a minimum of 53%) and an average recall of 34% (with a maximum of 59% and a minimum of 15%).

V. DISCUSSION

Based on the results shown in Section IV-E, we conducted a detailed analysis using the 07:33 departure of the Industrial Technology Center Route. This route had the highest number of passengers and yielded intermediate values for both precision and recall, making it suitable for examining both the effectiveness and limitations of the proposed method.

Fig. 7 visualizes the comparison between the ground truth and estimated OD data. Red segments indicate ground truth OD data, while blue segments show the OD data estimated by our method. The total number of estimated passengers was 18, of which 11 were correctly identified. Also, we visualized the data for iPhone devices after applying the address carry-over process, as shown in Fig. 6. The x-axis represents time, and the y-axis shows different devices. Horizontally aligned entries are MAC addresses presumed to belong to the same device. Different colors in the lines indicate different MAC addresses. Red dashed lines represent bus stop arrival times, and blue dashed lines represent departure times.

By analyzing these figures, we explored the potential for improving recall and precision. Labels indicating successful or failed estimations are placed to the left of each trace. Below, we detail each label and its implication:

- Red circles (◯) represent correctly estimated OD pairs.
- Crosses (×) represent incorrect estimations.
- Triangles (△) indicate overestimations due to users carrying multiple devices.
- **TH-1** represents failures due to MAC address changes right before alighting. The new MAC address appears only briefly and is filtered out by the appearance count threshold. If the threshold is too low, noise increases (lower precision); if too high, valid but short-lived MAC addresses are missed (lower recall).
- **TH-2** indicates cases where MAC addresses are filtered out because their appearance/disappearance falls outside the ± 30 second boarding/alighting detection window. Expanding the window increases recall but risks false positives, lowering precision. Shrinking the window has the opposite effect: precision improves, but recall may drop, especially when users are near the bus post-alighting.
- **a-on** indicates that the MAC address appeared near a boarding stop with a known passenger, but no corresponding alighting passenger exists near its disappearance time.
- **a-off** is the reverse of a-on that the MAC address disappears near a valid alighting stop, but no boarding passenger is matched to its appearance time.

From the failure of **TH-1** and **TH-2**, considering these trade-offs, the current time-based detection method has limitations. To improve performance, more robust approaches incorporating additional features such as RSSI trends should be explored in future work.

Also, the failure of **a-on** and **a-off** (one-sided matches) suggests that many MAC addresses partially overlap with real passengers but lack full OD correspondence. This may be due to noise or incomplete tracking due to address randomization.

VI. CONCLUSION

In this study, we proposed and evaluated a method for estimating Origin-Destination (OD) data on route buses using BLE advertising packets. Traditional approaches face challenges such as high costs and the difficulty of tracking due to randomized MAC addresses. To address these issues, our proposed method implements a carry-over mechanism that leverages the asynchronous nature of MAC address changes and information embedded in BLE packets. This enables continuous tracking of devices and effective estimation of OD data. We conducted evaluations on two distinct bus routes in Okayama City. The results showed that the proposed method achieved an average precision of 70% and an average recall of 34%. Future work includes a detailed investigation of the conditions under which iPhones and Android smartphones emit BLE advertising packets, dynamic optimization of thresholds for address carry-over and appearance counts, improving boarding/alighting detection using RSSI values, and supplementing BLE-based tracking with Wi-Fi to include devices that do not emit BLE packets. Additionally, the use of dedicated BLE sniffing devices with enhanced sensitivity and reception capabilities could potentially capture a greater number of advertising packets compared to standard USB Bluetooth adapters, thereby reducing packet loss and improving the overall recall performance of the system. These enhancements could significantly improve the accuracy of OD data estimation.

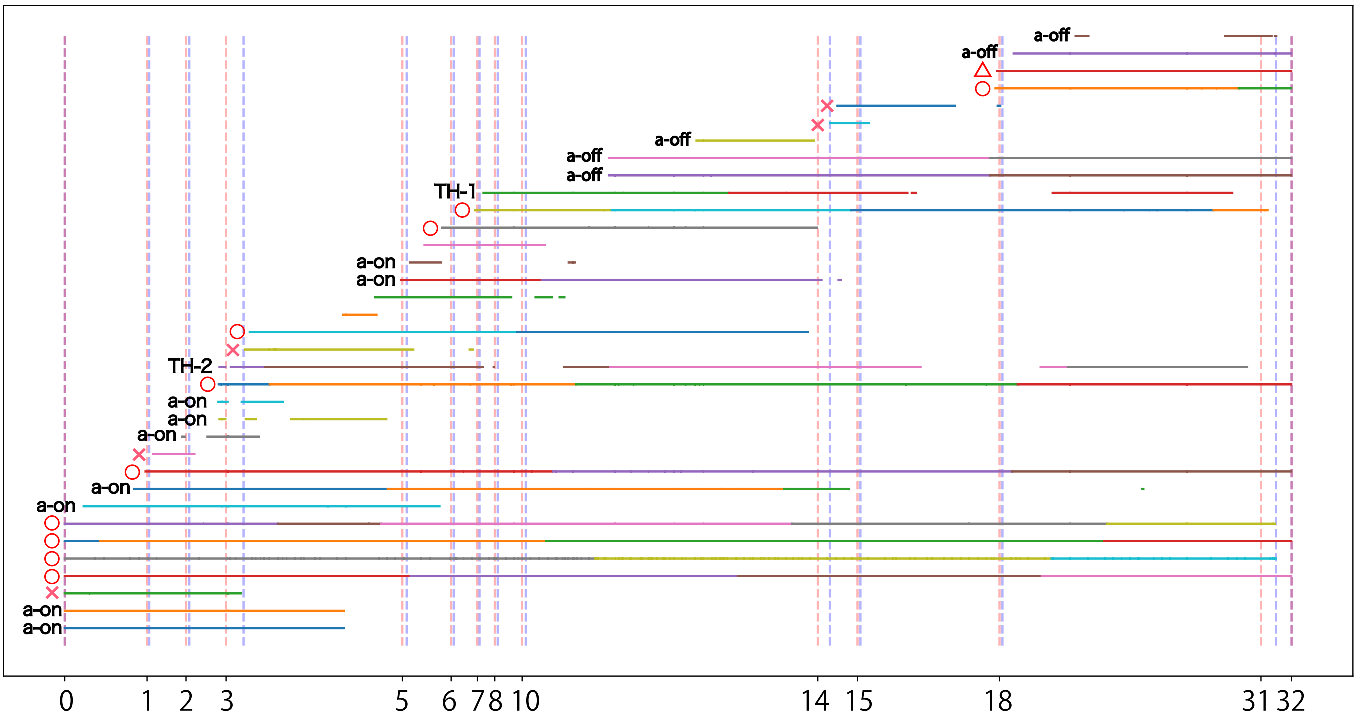

Fig. 6. Graph after address carry-over (iPhone)

Alighting bus stop / Boarding bus stop	2		3		7		14		15		18		31		32	
	T	E	T	E	T	E	T	E	T	E	T	E	T	E	T	E
0 Bus stop A	0	0	0	1	0	0	1	0	0	0	0	0	3	2	5	2
1 Bus stop B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2 Bus stop C			0	0	0	0	0	0	0	0	0	0	1	0	2	0
3 Bus stop D					0	1	7	1	1	0	0	0	2	0	2	1
5 Bus stop E					0	0	2	0	0	0	0	0	1	0	0	0
6 Bus stop F					0	0	1	1	0	0	0	0	1	0	0	0
7 Bus stop G							0	0	0	0	0	0	1	1	0	0
8 Bus stop H							0	1	0	0	0	0	1	1	0	0
10 Bus stop I							0	0	0	0	0	0	0	0	1	0
14 Bus stop J									0	1	0	1	0	0	0	0
15 Bus stop K											0	0	0	0	0	0
18 Bus stop L													1	0	1	2
31 Bus stop M															0	0
32 Bus stop N																

Fig. 7. Result of OD estimation (T: True OD, E: Estimated OD)

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan, "Current Status and Issues Surrounding Regional Public Transportation and Perspectives for Consideration," https://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000211.html, 2023, accessed on 06/06/2025 (in Japanese).

[2] Z. Wang, S. Wang, and H. Lian, "A route-planning method for long-distance commuter express bus service based on od estimation from mobile phone location data: The case of the changping corridor in Beijing," *Public Transport*, vol. 13, pp. 101–125, 2021.

[3] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan, "Encyclopedia of bus data application," https://wwwtb.mlit.go.jp/chubu/tsukuro/library/, 2020, accessed on 06/06/2025 (in Japanese).

[4] W. Liu, Q. Tan, and L. Liu, "Destination estimation for bus passengers based on data fusion," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 8305475, 2020.

[5] I. Lee, S.-H. Cho, K. Kim, S.-Y. Kho, and D.-K. Kim, "Travel pattern-based bus trip origin-destination estimation using smart card data," *PLOS ONE*, vol. 17, no. 6, pp. 1–20, 06 2022.

[6] C. Zhang, X. Chen, J. Zhao, Z. Jiang, and E. Chung, "Estimating bus passenger origin-destination flow via passenger reidentification using video images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 6, pp. 8353–8367, 2025.

[7] S. Ryu, B. B. Park, and S. E.-T. Ryu, "WiFi Sensing System for Monitoring Public Transportation Ridership: A Case Study," *KSCE Journal of Civil Engineering*, vol. 24, pp. 3092–3104, 2020.

[8] M. Kawashima, I. Arai, A. Endo, M. Kakiuchi, and K. Fujikawa, "Origin Destination Estimation Carrying Over Rolling Proximity Identifiers with RSSI," in *2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, 2022, pp. 7–12.

[9] Z. Pu, M. Zhu, W. Li, Z. Cui, X. Guo, and Y. Wang, "Monitoring Public Transit Ridership Flow by Passively Sensing Wi-Fi and Bluetooth Mobile Devices," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 474–486, 2021.

[10] Digital Agency, "New Coronavirus Contact Confirmation Application (COCOA)," https://www.digital.go.jp/policies/cocoa/, 2020, accessed on 06/06/2025 (in Japanese).

[11] S. Akiyama and Y. Taniguchi, "Improvement and evaluation of a method for estimating same device by BLE packets from moving devices with randomized MAC addresses," in *Proceedings of 2022 Kansai Branch Convention of IPSJ*, vol. 2022, no. G-46, 2022, pp. 1–4.

[12] J. K. Becker, D. Li, and D. Starobinski, "Tracking Anonymized Bluetooth Devices," *Privacy Enhancing Technologies Symposium (PETS)*, vol. 2019, pp. 50–65, 2019.

[13] T. Yoshimura, K. Hayashi, I. Arai, and Y. Matsuda, "Initial Study on OD Data Estimation Using BLE Random MAC Addresses," in *Proceedings of 2024 Kansai Branch Convention of IPSJ*, no. G-11, 2024, pp. 1–8.

[14] Y. Matsuda, H. Suwa, K. Hayashi, T. Yoshimura, A. Yoshihara, and I. Arai, "Estimating People Flow and Crowdedness for Various Urban Environments based on BLE Signal Sensing: Practical Studies," pp. 1–11, 2025. [Online]. Available: https://doi.org/10.23919/transcom.2025CEI0006