People Flow Estimation in Urban Environments Using BLE Advertising Packets

Kotaro Hayashi*[1], Taito Yoshimura[2], Hiroyuki Kobayashi[3], Ryo Imamura[4], Mina Itou[5], Kiyohide Ushijima[6], Hirohiko Suwa[7], Yuki Matsuda[8]
[1,2,8]Okayama University, [3,5,6]Local Media Labs Inc., [4]Saga City Office, [7,8]Nara Institute of Science and Technology, [7,8]RIKEN AIP

Abstract

In recent years, the analysis of people flow has become increasingly important for urban management and commercial strategies. While cameras and 3D sensors have been used for people flow estimation, they involve high costs and raise privacy concerns. In contrast, Wi-Fi and Bluetooth signals enable low-cost, non-intrusive data collection. However, widespread MAC address randomization in modern devices limits the effectiveness of conventional people tracking methods that rely on static MAC addresses. To overcome this limitation, we propose a method to estimate movement trajectories by associating dynamically randomized MAC addresses. Our approach utilizes the appearance/disappearance times and advertising (Adv) data of BLE packets collected by multiple scanners. Unlike prior work that focused on single-sensor setups or simulations, we conducted a citywide field experiment spanning 2.5 km in central Saga City, Japan. Thirty BLE scanners were deployed over a 60-day period, and the collected data were evaluated against GPS logs as the ground truth. The proposed method achieved an average macro coverage of 70.9% for walking, 46.2% for cycling, and 66.0% overall. This represents improvements of 9.5%, 3.4%, and 8.3% over baseline methods, respectively.

1 Introduction

Estimating people flow in urban spaces is essential for effective city management and planning of business strategies. According to Japan's Ministry of Land, Infrastructure, Transport and Tourism (MLIT) [1], people

[1]kotaro.hayashi@cocolab.jp
[8]yukimat@okayama-u.ac.jp

flow data supports three key applications: understanding the current conditions, evaluating policy effectiveness, and serving as a basis for forecasting. When integrated with other datasets, such data enables evidence-based policy making (EBPM), leading to growing demand for accurate and scalable flow estimation.

Manual observation of pedestrian movements requires considerable labor and cannot capture detailed trajectories across city-scale areas. To address this issue, vision-based approaches using cameras and 3D sensors have been proposed [2, 3, 4, 5]. However, these methods entail high deployment and operating costs and raise privacy concerns. Alternative techniques leveraging Wi-Fi or Bluetooth packet data [6, 7, 8] are more scalable, but their effectiveness has declined due to MAC address randomization in modern devices.

To overcome these limitations, we propose a method to estimate people flow by associating dynamically randomized MAC addresses captured by multiple BLE scanners. The association is based on the timing and content of appearance and disappearance events in BLE advertising packets. To evaluate the proposed method, we deployed 30 BLE scanners along a 2.5 km area centered around Saga Station, Japan, for 60 days. We also collected GPS logs as ground truth and quantitatively compared the estimated trajectories with the actual ones. The proposed method achieved an average trajectory coverage of 70.9% for walking, 46.2% for cycling, and 66.0% overall—representing 9.5%, 3.4%, and 8.3% improvements, respectively, over the unassociated baseline—demonstrating its effectiveness.

2 Related Work

2.1 People Flow Estimation

Various approaches have been proposed for estimating people flow using cameras and sensors. Wang *et al.* and Xie *et al.* developed multi-camera person re-identification (Re-ID) methods for tracking pedestrian movements [2, 3]. Nagata *et al.* estimated people flow using 3D sensors installed at entrances and exits, leveraging physical attributes such as height, walking speed, and transit time [4]. Yamaguchi *et al.* proposed a LiDAR-based system, "Hitonavi," which tracks pedestrian movements without capturing images [5]. While these methods effectively monitor individual trajectories, they involve high installation and maintenance costs and often raise privacy concerns. This motivates the need for lower-cost and privacy-preserving alternatives.

Consequently, with the widespread use of smartphones, Wi-Fi and Bluetooth-based crowd sensing has become a major research focus. Matsuda *et al.* estimated congestion levels in restaurants and public facilities using BLE signals [9], while Goto *et al.* analyzed bidirectional pedestrian flow by clustering BLE time-series differentials [10]. These studies effectively capture localized congestion but do not reconstruct trajectories.

Other research has applied Wi-Fi and BLE sensing to large-scale environments. Weppner *et al.* monitored event congestion using mobile device signals [11], and Versichele *et al.* proposed Bluetooth proximity

tracking using multiple scanners [6]. Mochiduki *et al.* reconstructed device trajectories by linking detections across multiple Wi-Fi scanners [7], and Traunmueller *et al.* analyzed urban mobility in New York using Wi-Fi probe data [8]. All of these methods rely on fixed MAC addresses as device identifiers. However, since Bluetooth v4.0 introduced private address randomization[2]—now standard in iOS[3,4] and Android[5,6] —the practicality of fixed-address approaches has declined, thereby creating demand for identification-free tracking techniques.

2.2 Association of Randomized MAC Addresses

The adoption of randomized MAC addresses in BLE and Wi-Fi improves privacy but complicates continuous device tracking. To address this, several studies have explored methods for associating randomized addresses. Becker *et al.* proposed an address carryover algorithm linking BLE addresses via identifiers in advertising packets [12], although this vulnerability has since been patched. Jouans *et al.* treated transmission intervals as weak identifiers to associate devices [13], and Boussad *et al.* generated RSSI-based symbolic feature vectors for association [14]. Akiyama *et al.* applied regression-based analysis to link pseudonymized addresses [15], while Mishra *et al.* used logistic regression on Wi-Fi probe request features and validated their model with pre-randomization and Hong Kong datasets [16].

Although these studies address address-association challenges, most rely on static or simulated environments and lack validation under real-world conditions. Applying such methods to dynamic, people-flow estimation contexts remains an open research challenge.

2.3 Position of this study

In contrast to conventional people flow estimation methods using high-cost sensors such as cameras or LiDAR, the proposed approach employs BLE scanners to achieve low-cost and privacy-preserving sensing. It can capture movement information without identifying individuals or recording physical features, providing a non-invasive and contactless solution. Beyond aggregate-level sensing, our method links dynamically randomized MAC addresses to track individual devices without relying on fixed identifiers, offering a more realistic framework for practical deployment.

A key distinction of this study is its validation under real-world urban conditions. Whereas prior work on MAC address association has often relied on indoor or synthetic environments, we deployed multiple BLE scanners throughout an actual city area for empirical evaluation. Our previous preliminary work [17] examined the feasibility in a limited campus corridor, but did not assess scalability to urban contexts. In this

[2] `https://www.bluetooth.com/specifications/specs/core-specification-4-0`
[3] `https://support.apple.com/en-hk/guide/security/sec82597d97e/web`
[4] `https://support.apple.com/en-hk/guide/security/secb9cb3140c/web`
[5] `https://source.android.com/docs/core/connect/bluetooth`
[6] `https://source.android.com/docs/core/connect/wifi-mac-randomization-behavior`

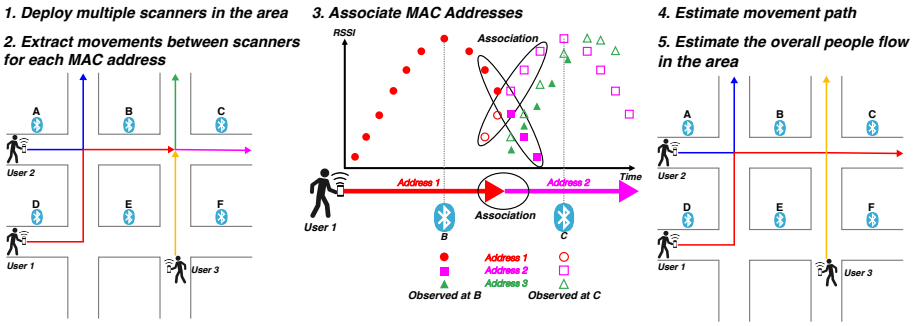

Figure 1: Overview of the Proposed Method

study, we conducted large-scale outdoor experiments and quantitatively evaluated tracking accuracy using GPS-based ground truth.

In summary, this paper advances BLE-based people flow estimation by emphasizing six aspects: low cost, privacy preservation, support for randomized MAC addresses, individual-level tracking, real-world deployment, and objective accuracy validation.

3 Proposed Method

An overview of the proposed method is shown in Figure 1, and the details of its five steps are described below. The specific processing steps are also presented as pseudocode in Algorithm 1.

1. **Deploy multiple scanners in the area** Multiple BLE scanners are deployed throughout the experimental area to collect data from BLE devices. BLE Adv packets contain information such as manufacturer-specific data (MSD) and 16-bit UUIDs that can identify the manufacturer of the transmitting device. Therefore, the data is analyzed separately for each manufacturer (*filter_by_manufacturer* function in Algorithm 1). For iPhones, most devices transmit packets whose MSD begins with `0x4c0010`; therefore, only packets starting with this value are targeted. For Android devices, the analysis targets packets containing the 16-bit UUID `0xfef3`, which is commonly observed and assigned to Google LLC by the Bluetooth SIG.

2. **Extract movements between scanners for each MAC address** Observations of the same MAC address from different scanners are integrated (*aggregate_by_address* function in Algorithm 1). First, identical MAC addresses observed across scanners are extracted to generate tuples consisting of scanner IDs and timestamp. These tuples are then sorted chronologically. If consecutive observations occur at the same scanner, the timestamps of the first and last observations are retained. This process results in data like: "id1 10:00–10:03, id2 10:05–10:08, id1 10:10–10:30...". This format allows the estimation of the dwell time and order of movement at each scanner before the device randomizes its MAC address. In

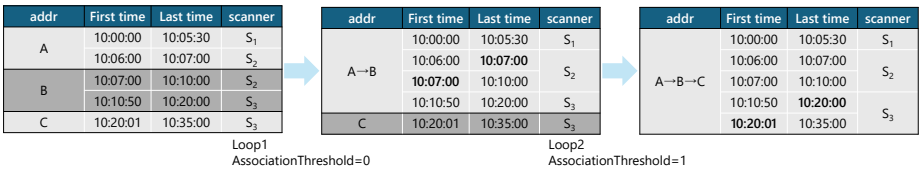

Figure 2: Worked Example of Address Association

Figure 1, the arrows of the same color in step 2 indicate the same MAC address. For example, in the case of red arrows, the device is first detected at scanner D and then at scanner B, indicating movement from D to B.

3. **Associate MAC Addresses** MAC addresses observed only at a single scanner and whose disappearance-to-appearance time difference is below a certain threshold (`ObservationTimeThreshold`) are considered to have little value for movement estimation and are filtered out (*denoise_by_time_and_move* function). For the remaining MAC addresses, association is performed when three conditions are satisfied. First, the time difference between disappearance and reappearance is minimal; second, the absolute time gap is within a predefined threshold (*AssociationThreshold*); and third, the scanner ID before disappearance matches the scanner ID immediately after reappearance at the same location with a new randomized MAC address (*associate_address* function). Figure 2 shows a worked example of the association process, illustrating how randomized MAC addresses are merged into a single device trajectory.
4. **Estimate movement path** Step 3 is recursively repeated until no more address pairs satisfying the association conditions remain. This process enables the estimation of the trajectory of a single BLE device within the experimental area (*associate_address* function).
5. **Estimate the overall people flow in the area** By applying Step 4 to all BLE devices, the proposed method estimates the overall people flow within the experimental area (*associate_address* function).

4 Experiment and Evaluation

4.1 Experiment Overview

The experiment was conducted over a 60-day period from September 26 to November 24, 2024. The experimental area, shown in Figure 3, spanned approximately 2.5 km in Saga City, Saga Prefecture, extending from “SAGA ARENA” through “Saga Station” to “Shirayama Shopping Street.” A total of 30 BLE scanners were installed along this route, and each scanner was assigned a unique scanner ID from 1 to 30. To evaluate the performance of the proposed method, ground-truth data were generated by processing GPS data to serve as the evaluation baseline. GPS data were collected by participants carrying both a GPS logger and

Algorithm 1: Proposed Method

```
Function filter_by_manufacturer(scan_data, target_manufacturer):
    filtered_array ← [ ];
    forall row ← scan_data do
        if row[ADdata][MSD] begins with '4c0010' and target_manufacturer is 'apple' then
            filtered_array.append(row);                                  // apple device
        else if row[ADdata][16bitUUID] is 'fef3' and target_manufacturer is 'google' then
            filtered_array.append(row);                                  // google device
    return filtered_array;

// Function to aggregate scan data from multiple scanners
Function aggregate_by_address(filtered_array):
    aggregated_array ← [ ];
    forall target_addr ← unique(filtered_array[addr]) do
        target_addr_arr ← [ ];
        forall data ← filtered_array do
            if data[addr] is target_addr then
                target_addr_arr.append(data);                            // get target_addr data
        scanner_intervals ← ExtractContinuousScannerIntervals(target_addr_arr);
        // extract continuous stay intervals segmented by scanner_id
        // each interval is represented as (first_timestamp, last_timestamp, scanner_id)
        aggregated_array.append({addr : target_addr, intervals : scanner_intervals});
    return aggregated_array;

Function denoise_by_time_and_move(aggregated_array):
    denoised_array ← [ ];
    ObservationTimeThreshold ← 10;                                   // filtering threshold (minutes)
    forall data ← aggregated_array do
        if (GetLastTimestamp(data) − GetFirstTimestamp(data) ≥
          ObservationTimeThreshold) or GetCountScanner(data) ≥ 2 then
            // If observed duration exceeds TimeThreshold or observed by multiple scanners
            denoised_array.append(data);
    return denoised_array;

Function associate_address(denoised_array):
    AssociationThreshold ← 10;                                                          // seconds
    // threshold of time difference between appearance and disappearance time for association
    candidate_group ← [ ];
    forall record ← denoised_array do
        candidate_group.append([record]);
    forall i ← range(0, AssociationThreshold) do
        associated_array ← [ ];
        while len(candidate_group) > 0 do
            associated_data ← [ ];
            end_at, last_scanner ← GetLastTimestampAndScanner(candidate_group[0]);
            associated_data.append(candidate_group.pop(0));
            found ← true;
            while found do
                found ← false;
                forall candidate ← candidate_group do
                    if (abs(GetFirstTimestamp(candidate) − end_at) ≤ i) and
                      last_scanner is GetFirstScanner(candidate) then
                        // If |appearance time - disappearance time| ≤ i and the scanner
                           after appearance is the same as the scanner before disappearance
                        end_at, last_scanner ←
                          GetLastTimestampAndScanner(candidate);
                        associated_data.append(candidate);
                        candidate_group.remove(candidate);               // candidate is unique
                        found ← true;
                        break;                                           // break forall loop
            associated_array.append(associated_data);
        candidate_group ← associated_array;
    return associated_array;

Function main(scan_data):
    scan_data ← AllScannedBleAdvPackets;
    // AllScannedBleAdvPackets:  array of all data observed by all scanners
    target_manufacturer ← 'apple';                  // target_manufacturer = 'apple' or 'google'
    filtered_array ← filter_by_manufacturer(scan_data, target_manufacturer);
    aggregated_array ← aggregate_by_address(filtered_array);
    denoised_array ← denoise_by_time_and_move(aggregated_array);
    return associate_address(denoised_array);
```

Figure 3: Scanner Deployment Map and Photos

a BLE-emitting device (iPhone). These participants moved within the experimental area during data collection. The ground-truth dataset consisted of 15 samples: 12 walking patterns and 3 cycling patterns. This study was approved by the Ethical Review Committee for Research Involving Human Subjects at Okayama University Shizen-2024-12 and Nara Institute of Science and Technology 2020-I16.

4.2 Scanner Configuration

The BLE Adv packet scanners were built using either the Raspberry Pi 4 Model B[7] or the Raspberry Pi 5[8], as shown in Figure 4. For BLE data collection, a USB adapter compatible with Bluetooth 4.0+EDR/LE Class 1 (BUFFALO, model BSBT4D100) was attached. Although the Raspberry Pi 4 Model B and Raspberry Pi 5 have built-in Bluetooth communication modules, an external module was used to ensure consistent Bluetooth specifications across all scanners. The bluepy library for Python[9] was used to collect Adv packets. Additionally, LTE communication modules

[7] https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
[8] https://www.raspberrypi.com/products/raspberry-pi-5/
[9] https://github.com/IanHarvey/bluepy

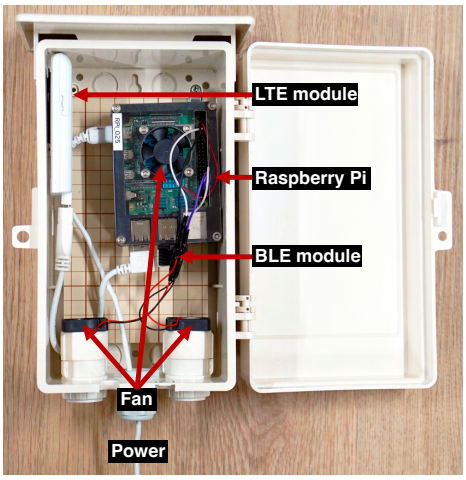

Figure 4: Scanner Configuration

were used for remote monitoring of scanner operation and time synchronization. Specifically, an LTE-compatible USB dongle (PIXELA, model PIX-MT110, or FUJISOFT, model +F FS040U) was connected. The experiment started in late September, when temperatures in Japan can still exceed 30°C due to its humid climate. Therefore, we implemented measures to prevent scanner overheating. Specifically, fans were installed in both the scanner unit and the plastic weatherproof box, and a heat sink was attached to the external Bluetooth communication module to enhance heat dissipation. To protect the device from rain and wind, a plastic weatherproof box (manufactured by Mirai Industry, model WB-1AM) was used.

5 Results and Discussion

During the 60-day experimental period, a total of approximately 900 million BLE Adv packets were collected across all scanners, corresponding to about 18 million unique MAC addresses. Additionally, we collected 15 GPS-based movement logs: 12 walking samples and 3 cycling samples.

5.1 Evaluation Method

First, we applied the proposed method to all BLE data collected during the experimental period, after excluding the data from Scanner 1, which lacked time synchronization. This process generated the estimated trajectory data. The filtering threshold ($ObservationTimeThreshold$) used in the proposed method was set to 10 minutes. This value was chosen based on the Bluetooth Core Specification (versions 4.0[2] to 5.1[10]), which recommends a minimum interval of 15 minutes for private address changes.

[10] https://www.bluetooth.com/specifications/specs/core-specification-amended-5-1

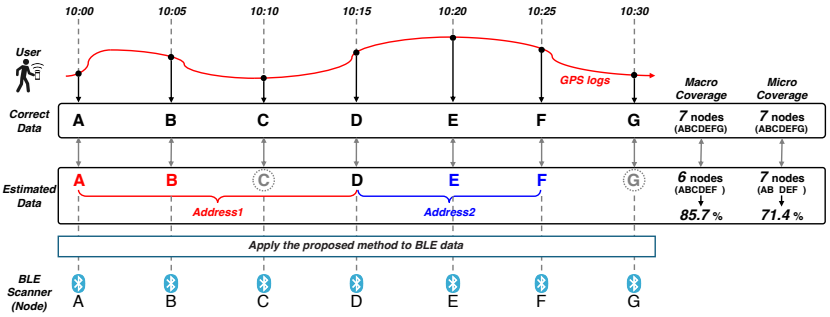

Figure 5: Evaluation Metrics

In addition, the association threshold (*AssociationThreshold*) was set to 10 seconds to account for possible missed packets. Next, the estimated trajectory data obtained using the proposed method were compared with the ground-truth data generated from GPS logs recorded while the subject carried both a GPS logger and a BLE-transmitting device (iPhone). To assess the results, two evaluation metrics were defined (Figure 5). In both evaluation metrics, if the estimated trajectory data were fragmented into multiple segments, only the largest fragment was used for calculation.

Macro Coverage The proportion of the area covered by the estimated trajectory data relative to the ground-truth data. For example, in Figure 5, out of the seven nodes from A to G, the estimated trajectory covers 6 nodes (A to F). Although scanner C failed to capture the address, the trajectory from A to F was successfully estimated, so the missing node is considered to be interpolated and thus covered.

Micro Coverage The proportion of nodes passed in the ground truth that are also included in the estimated trajectory (strict coverage rate). In Figure 5, out of the seven nodes from A to G, only 5 nodes (A, B, D, E, and F) are covered. Since scanner C failed to capture the address and no interpolation was made, it was considered not covered.

The ground-truth data consisted of timestamps extracted at the moments when the subject passed in front of each scanner, with each scanner treated as a node. Each entry in this data was a pair consisting of a scanner ID and its corresponding timestamp.

In this analysis, we first identified the MAC address corresponding to the iPhone carried by the subject with the GPS logger from the BLE data, taking into account the timestamps and the order of scanner visits in the ground-truth data.

5.2 Comparison with GPS data

Subsequently, we compared the estimated trajectory data obtained using the proposed method with the ground-truth data, calculated the macro

Table 1: Summary of Evaluation Results for the Proposed Method

Route ID	Transport Mode	Area	Correct Nodes	Before Matching (Macro)		Associated Addresses	After Matching (Macro)		After Matching (Micro)	
				Nodes	Coverage		Nodes	Coverage	Nodes	Coverage
1	Walking	1-30	42	18	42.9%	2	30	71.4%	30	71.4%
2	Walking	4-9	4	3	75.0%	1	4	100.0%	4	100.0%
3	Walking	1-10	7	6	85.7%	0	6	85.7%	6	85.7%
4	Walking	1-10	13	5	38.5%	0	5	38.5%	5	38.5%
5	Walking	1-10	13	7	53.8%	1	9	69.2%	9	69.2%
6	Walking	7-30	45	20	44.4%	1	26	57.8%	24	53.3%
7	Walking	6-21	12	9	75.0%	0	9	75.0%	7	58.3%
8	Walking	15-21	6	3	50.0%	0	3	50.0%	3	50.0%
9	Walking	7-30	42	17	40.5%	1	25	59.5%	23	54.8%
10	Walking	20-30	16	8	50.0%	0	8	50.0%	7	43.8%
11	Walking	20-30	16	13	81.3%	1	15	93.8%	15	93.8%
12	Walking	21-23	5	5	100.0%	0	5	100.0%	5	100.0%
13	Cycling	1-9	6	3	50.0%	0	3	50.0%	3	50.0%
14	Cycling	20-30	17	4	23.5%	0	4	23.5%	3	17.6%
15	Cycling	17-30	20	11	55.0%	1	13	65.0%	13	65.0%
	Avg. (Walking)				61.4%			70.9%		68.2%
	Avg. (Cycling)				42.8%			46.2%		44.2%
	Avg. (Overall)				57.7%			66.0%		63.4%

coverage and micro coverage values as shown in Table 1. The results indicate that, when comparing walking and cycling patterns, the walking trajectories achieved higher macro and micro coverage. This is attributed to the inherent difficulty of scanning BLE packets from fast-moving devices, such as those carried by users while cycling. In the estimated trajectory data, two major types of address association failures were observed as described below.

5.2.1 Failures Caused by Physical Distance Between Scanners

We examined the causes of MAC address association failures due to the physical distance between scanners, taking Route ID 1 and Route ID 15 as illustrative examples.

As shown in Figure 6 (Route ID 1, walking pattern), the device address failed to be associated twice because of the long distance between scanners. The first failure is considered to have occurred because Scanner 5 was out of operation, resulting in a large gap of approximately 500 m between adjacent scanners. When the distance between scanners increases, there are intervals during which no scanner can capture the device address; consequently, the address may change during that period, leading to association failure. The second failure appears to have occurred because the address could not be captured by Scanners 15, 14, 13, and 12. As shown in Figure 6, the address was not captured by Scanner 15 because the user was walking on the opposite side of the roadway from the scanner's installation location. The BLE advertising packets transmitted by the user's device may not have reached the scanner due to obstacles such as vehicles or walls beside the scanner. Furthermore, Scanners 12, 13, and 14 operated using the power supply for the illumination system, and they were inactive when the user passed in front of them because it was outside the illumination period. As a result, the effective distance between scanners increased to approximately 400 m, which likely caused the failure of association.

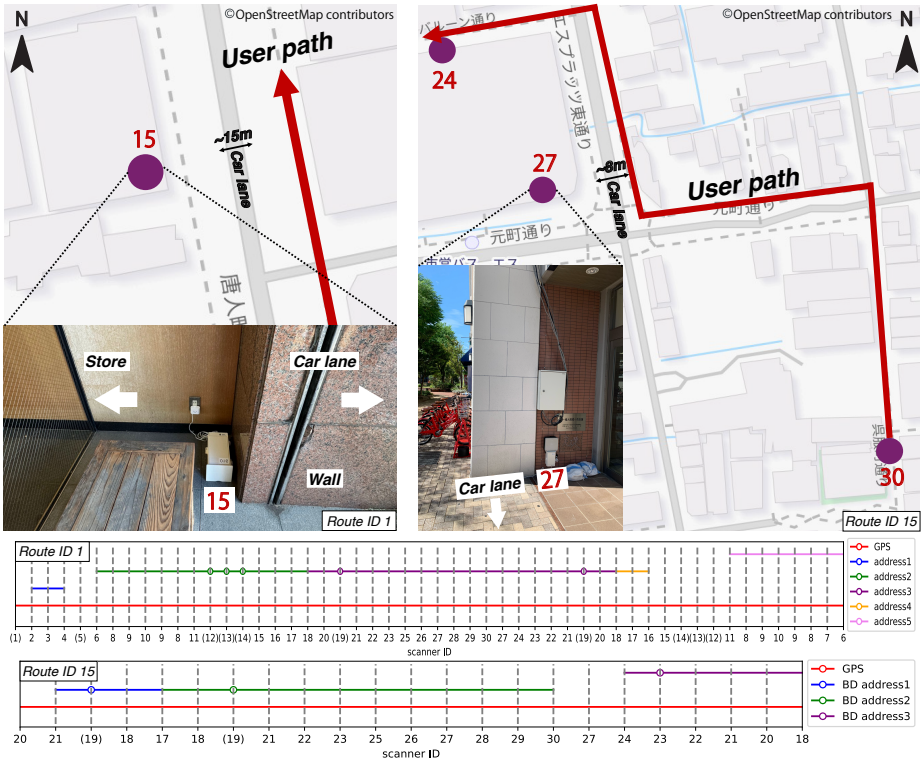

Figure 6: Visualization of Route IDs 1 and 15, and Locations of Scanners 15 and 27

As shown in Figure 6 (Route ID 15, Cycling pattern), the device address failed to be associated once because of the physical distance between scanners. The failure to capture the address at Scanner 27 likely occurred because the user was traveling on the opposite side of the roadway from the scanner's installation location, and the use of a bicycle may also have contributed to the result. The BLE advertising packets transmitted from the user's device may not have reached the scanner due to obstacles such as vehicles, or the user's high traveling speed may have prevented the scanner from detecting the signal. This association failure is similar in mechanism to that observed for Route ID 1. Because the address was not captured by Scanner 27, the effective distance between adjacent scanners increased to approximately 300 m, during which the address may have changed, leading to the association failure.

Such association failures caused by the physical distance between scanners were also observed in five additional routes. In total, nine association failures were recorded across seven of the fifteen ground-truth GPS routes (an overview is shown in Table 2). Although the maximum communication range of Bluetooth varies depending on the device, it is generally estimated to be approximately 100 m. Therefore, when the distance be-

Table 2: Summary of Detection Failures Due to Physical Distance

Scanner Pair	Distance	Number of Cases
4-6	Approx. 500m	4
11-16	Approx. 400m	1
17-18	Approx. 300m	1
30-24	Approx. 300m	1
3-4	Approx. 200m	1
30-27	Approx. 200m	1

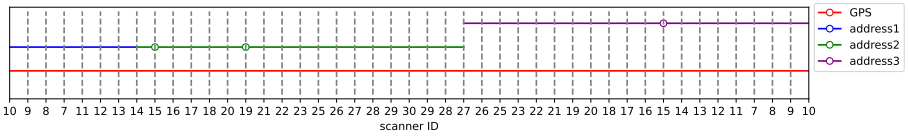

Figure 7: Route ID 6

tween scanners exceeds 200 m, a time interval occurs during which device addresses cannot be captured, thereby increasing the likelihood of association failure. This tendency was also confirmed in real-world environments. These results suggest that reducing the spacing between scanners and deploying them at a higher density could lower the probability of address-capture failures. However, because the installation and maintenance costs of scanners increase proportionally, further experiments are required to determine the optimal scanner spacing that balances cost and performance.

5.2.2 Failures Caused by Incorrect Address Association

Figure 7 illustrates a case of Route ID 6, in which address association failed because a device was mistakenly associated with an incorrect MAC address. Both address2 and address3 were observed by the same scanner; therefore, a correct association between them was expected. However, address2 was erroneously associated with another address (addressX), failing the intended association. In the proposed method, addresses were filtered using the MSD, and subsequently associated based solely on their appearance and disappearance times. As a result, the simultaneous appearance of addressX and address3 is considered to have caused the incorrect association. Additionally, two other cases were identified in which addresses clearly not belonging to user devices were incorrectly associated, although these did not directly contribute to other association failures. These incorrect associations are presumed to have occurred because the association process relied exclusively on temporal information. The proposed method is based on the assumption that devices from the same manufacturer are unlikely to simultaneously change their MAC addresses at the same location and time. However, the manufacturer filter cannot distinguish between different device types from the same manufacturer (e.g., iPhone and iPad, or MacBook). Therefore, when multiple devices from the same manufacturer are present in close proximity, incorrect associations may occur. This risk is expected to increase in higher-density environments

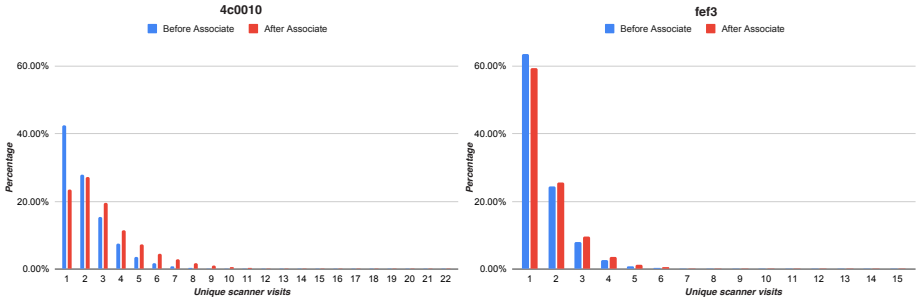

Figure 8: Proportion of Scanner Visits: Before vs. After Association

where a larger number of devices from the same manufacturer coexist. In future work, incorporating additional features such as RSSI variation patterns, packet transmission intervals, and Tx Power Level, in addition to temporal information, could help mitigate such incorrect associations.

5.3 Comparison Before and After MAC Address Association

We compared the estimated trajectory data obtained by associating MAC addresses using the proposed method with the data prior to association. First, the macro coverage was calculated and summarized for both before and after applying the proposed method as shown in Table 1. The results show that the macro coverage increased by an average of 9.5% for walking and 3.4% for cycling after MAC address association. Figure 8 shows the proportion of scanner-visit counts before and after applying the proposed method for iPhone devices (packets with an MSD starting with `0x4c0010`) and Android devices (packets with a 16-bit UUID of `0xfef3`) throughout the entire experimental period. The figure demonstrates that, for both iPhone and Android devices, the proportion of devices with a larger number of scanner visits increased, whereas that of devices with fewer visits decreased after applying the proposed method. However, for Android devices, no trajectory data were obtained simultaneously with the GPS logger, preventing validation of the association accuracy against ground-truth data. In future work, we plan to extend the performance evaluation conducted with iPhones to Android devices as well, aiming to establish a method applicable to both platforms.

5.4 Visualization of Changes in People Flow

Finally, we discuss a visualization method for identifying changes in people flow between ordinary periods and event periods. Figure 9 illustrates the variation in people flow between ordinary periods and the period of the SAGA2024 National Sports Festival based on BLE advertising packets transmitted from Apple devices. Because BLE signals from Apple devices are generally more stable than those from Android devices, the proposed method was applied only to Apple-device packets in this analysis. The figure is oriented with north at the top and shows increases and

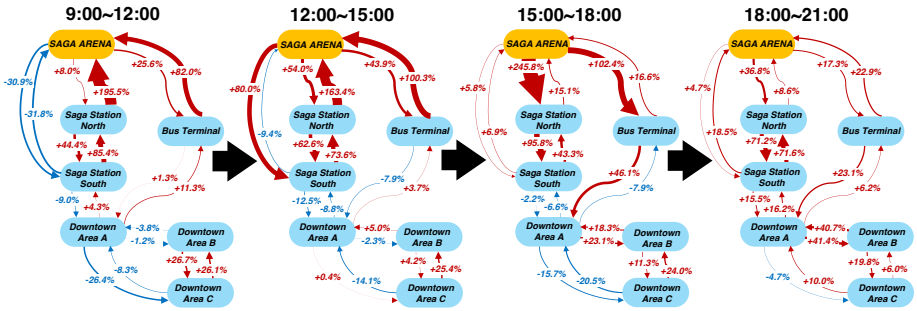

Figure 9: Change in People flow on Japan Games Holidays Compared to Regular Holidays

decreases in inter-area flow during the event period compared with ordinary periods for each time slot. From this visualization, it is evident that people flow between Saga Station and SAGA Arena increased throughout the day during the event period. In contrast, the people flow toward the downtown area—where Saga City aimed to attract more visitors and economic activity—showed only a modest increase at night. Based on this visualization, the Saga City provided the following feedback. People flow in the north–south direction connecting Saga Station and SAGA Arena increased substantially during the event, which was consistent with predictions derived from the spatial relationship between the main destination and the transportation hub. Conversely, for east–west movement within the city center, the aggregated data did not clearly capture continuous movement or significant event-induced changes. These findings underscore the importance of continuous, multi-perspective analyses of people flow to effectively promote pedestrian circulation within the city center.

5.5 Privacy and Security Considerations

Randomized MAC addresses were introduced to enhance user privacy and security, for example, by preventing continuous device tracking by third parties such as commercial entities. From a computer security perspective, MAC address association techniques could potentially weaken these protections. However, the proposed method enables only limited, area-specific tracking within a confined region where BLE scanners are densely deployed. Re-identification of individuals after they leave the monitored area and later re-enter it is not supported under our approach, as no persistent identifiers are maintained across spatially or temporally separated regions. This temporal and spatial constraint limits the tracking capability of the proposed method. In contrast to person re-identification using CCTV cameras, which can enable continuous cross-area tracking through facial or physical feature recognition, the proposed method detects only the presence of devices without capturing any biometric information, thereby offering a more privacy-preserving alternative to camera-

based tracking. The primary objective of this method is aggregate-level people flow estimation for urban planning and infrastructure evaluation, rather than individual tracking or identification. By focusing on collective mobility patterns at the city scale, the method provides valuable insights for transportation planning, facility management, and urban design while minimizing privacy risks associated with individual-level surveillance.

The proposed method relies on sufficiently dense deployment of BLE scanners to achieve reliable association performance, and its accuracy may degrade in sparse sensing environments. In addition, the current experimental validation was primarily conducted using iOS devices, and differences in BLE advertisement behavior across platforms may affect detection characteristics. Extending validation to other device types, including Android devices, is an important direction for future work. While camera- and LiDAR-based systems can provide more precise localization, they often involve higher deployment costs and privacy concerns. In contrast, the proposed BLE-based approach is intended as a complementary, low-cost, and privacy-aware alternative for aggregate-level people flow estimation.

Nevertheless, data security measures must be carefully implemented. In the event of a data breach, the collected MAC addresses could potentially be linked with external BLE datasets, unintentionally expanding the scope of tracking. To mitigate this risk, safeguards such as hashing or encrypting observed MAC addresses should be applied to prevent cross-dataset linkage, thereby reinforcing the spatial and temporal limitations of the tracking scope.

In summary, while MAC address randomization remains an essential privacy protection mechanism, the proposed method is designed with privacy considerations in mind. Future work should continue to explore technical and operational measures that further strengthen privacy protection while maintaining the utility of large-scale people flow analysis.

6 Conclusion

This paper proposes a people flow estimation method using randomized advertising packets, and demonstrates its effectiveness, limitations, and visualization results through a real-world experiment conducted in Saga City over 60 days. We confirmed that the proposed method showed an overall coverage of the estimated travel paths of 70.9% for walking, 46.2% for cycling, and 66.0% overall. Compared with the results before applying the method, the coverage improved by 9.5% for walking, 3.4% for cycling, and 8.3% overall, demonstrating the effectiveness of the proposed approach. The visualization results showed a clear increase in north–south flows between the Saga Station and the SAGA Arena during the "SAGA 2024 National Sports Festival," reflecting the relationship between major facilities and access points. These results suggest that the proposed method can achieve MAC address association and route estimation with a certain level of accuracy, and the results are also useful for city-wise people flow analysis. Future work includes experimental verification of optimal scanner placement intervals and improvement of the association algorithm to enhance matching accuracy.

Acknowledgments

This work was supported by Part of this research was supported by JSPS KAKENHI under Grant No. JP24K20763, JST PRESTO under Grant No. JPMJPR2039, JST COI-NEXT under Grant No. JPMJPF2115, and the SAGA Smart Machinaka Project Public-Private Partnership Demonstration Project. We would like to express our sincere gratitude to the local shop owners, public facilities, and other stakeholders in the experimental area for their kind cooperation in the installation of devices.

References

[1] The Ministry of Land, Infrastructure, Transport and Tourism. Guidelines for Utilizing People Flow Data to Solve Regional Issues Ver1.2 (in Japanese). `https://www.mlit.go.jp/tochi_fudousan_kensetsugyo/chirikukannjoho/tochi_fudousan_kensetsugyo_tk17_000001_00034.html`, 2024. Accessed on July 2, 2025.

[2] G. Wang, J. Lai, P. Huang, and X. Xie. Spatial-temporal person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8933–8940, 2019.

[3] Z. Xie, Z. Ni, W. Yang, Y. Zhang, Y. Chen, Y. Zhang, and X. Ma. A Robust Online Multi-Camera People Tracking System With Geometric Consistency and State-aware Re-ID Correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, CVPR '24, pages 7007–7016, June 2024.

[4] Y. Nagata, T. Yonezawa, and N. Kawaguchi. Person-Flow Estimation with Preserving Privacy Using Multiple 3D People Counters. In *Science and Technologies for Smart Cities*, pages 615–634, 2021.

[5] H. Yamaguchi, A. Hiromori, and T. Higashino. A Human Tracking and Sensing Platform for Enabling Smart City Applications. In *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking*, ICDCN '18, pages 1–6, 2018.

[6] M. Versichele, T. Neutens, M. Delafontaine, and N. Van de Weghe. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 32(2):208–220, 2012.

[7] M. Mochizuki, T. Onikura, Y. Fukuzaki, and N. Nishio. A Pedestrian Flow Analysis System using Wi-Fi Packet Sensors and its Application to a Real Environment. In *Proceedings of Multimedia, Distributed, Cooperative and Mobile Symposium*, DICOMO '14, pages 1249–1257, 2014.

[8] M. W. Traunmueller, N. Johnson, A. Malik, and C. E. Kontokosta. Digital footprints: Using WiFi probe and locational data to analyze

human mobility trajectories in cities. *Computers, Environment and Urban Systems*, 72:4–12, 2018.

[9] Y. Matsuda, K. Ueda, E. Taya, H. Suwa, and K. Yasumoto. BLECE: BLE-Based Crowdedness Estimation Method for Restaurants and Public Facilities. In *The 14th International Conference on Mobile Computing and Ubiquitous Networking*, ICMU '23, pages 1–6, 2023.

[10] I. Goto, K. Ueda, Y. Matsuda, H. Suwa, and K. Yasumoto. BLESS: BLE based Street Sensing for People Counting and Flow Direction Estimation. In *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events*, PerCom '24, pages 76–81, 2024.

[11] J. Weppner, B. Bischke, and P. Lukowicz. Monitoring Crowd Condition in Public Spaces by Tracking Mobile Consumer Devices with WiFi Interface. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 1363–1371, 2016.

[12] J. K. Becker, D. Li, and D. Starobinski. Tracking Anonymized Bluetooth Devices. In *Privacy Enhancing Technologies Symposium*, PETS '19, pages 50–65, 2019.

[13] L. Jouans, A. C. Viana, N. Achir, and A. Fladenmuller. Associating the Randomized Bluetooth MAC Addresses of a Device. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference*, CCNC '21, pages 1–6, 2021.

[14] Y. Boussad, Y. Yang, A. Tomlinson, and S. Grant-Muller. McMatcher: A symbolic representation for matching random BLE MAC addresses. In *2024 IEEE International Conference on Consumer Electronics*, ICCE '24, pages 1–6, 2024.

[15] S. Akiyama and Y. Taniguchi. A device identification method from BLE advertising packets with randomized MAC addresses based on regression of received signal strength. *IEICE Communications Express*, 13(3):64–67, 2024.

[16] A. K. Mishra, A. C. Viana, and N. Achir. Bleach: From WiFi probe-request signatures to MAC association. *Ad Hoc Networks*, 164, 2024.

[17] K. Hayashi, T. Yoshimura, H. Suwa, and Y. Matsuda. Exploration of City-wise People Flow Estimation Method Using Multiple BLE Scanners. In *Proceedings of 2024 Kansai Branch Convention of IPSJ*, number G-38, pages 1–8, 2024.