The Feasibility of Location Anonymity: An Empirical Study towards a Real-world Location Privacy Protection System in Takeout Services

Lu Zhou*¶, Ruoxu Yang†‡, Lichuan Ma*§, Guoxing Chen‡, Haojin Zhu‡, Li Yang*§, Qingqi Pei*§, Qiang Li¶
*State Key Laboratory of Integrated Services Networks, Xidian University
†Microsoft Asia-Pacific Technology Co., Ltd.
‡Shanghai Jiao Tong University
§Shaanxi key Laboratory of Blockchain and Secure Computing
¶Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University

***Abstract*—Various anonymity methods have been proposed to safeguard the privacy of human mobility trajectories, ranging from trajectory anonymity that uses a fixed pseudonym to location anonymity which involves using different pseudonyms for each location. While location anonymity appears to offer robust privacy protection, there is growing concern that trajectories can still be reconstructed even if this method is deployed. Due to the lack of evaluations on real-world systems utilizing location anonymity, its practical effectiveness remains uncertain.**

Two popular takeout platforms, Ele.me and Meituan, which have adopted location anonymity to protect riders' trajectories, provide a suitable environment for such real-world evaluations. We design a large-scale data collection system to gather anonymized location data of riders, creating two anonymized datasets containing millions of riders' locations. Then we propose an innovative multi-stage trajectory inference framework specifically tailored to location anonymity, containing location linking stage for short-term tracking and segment matching stage for long-term tracking. Extensive evaluations refute the effectiveness of location anonymity for short-term tracking (achieving 83.4% and 74.5% inference accuracy on Ele.me and Meituan) but confirm its utility for long-term tracking. Analysis highlights the crucial role of strong aggregation properties of riders, previously deemed unrealistic across multiple scenarios, in thwarting long-term tracking.

***Index Terms*—Location anonymity, Large-scale data collection, Multi-stage trajectory inference attack, Real-world evaluations**

I. Introduction

In recent decades, location-based services (LBSs) have become an essential component of numerous applications, including showing nearby users in mobile social networks [1], location tracking in fitness tracking social networks [2], and displaying nearby cars in ride-hailing services [3]. Typically, users' locations will be continuously or intermittently used. Frequent location privacy leakage events make users pay more attention to the usage of their location data, especially in continuous location services. The previous works have performed a comprehensive privacy analysis on different location-based services and pointed out that they suffer from a wide range of location privacy attacks, which can result in the loss of sensitive data such as the top locations (*e.g.*, office and home) and the moving trajectories [4], [5], [6], [7], [8].

A wide variety of location privacy-preserving approaches have been proposed to protect users' privacy [9]. Obfuscating locations to prevent the leakage of real locations is a key direction, such as location generalization (*e.g.*, cloaking) and location perturbation (*e.g.*, differential privacy) [10], [11], [12], [13], [14]. Over time, these techniques have evolved from manually setting the simple rules to adding random noises with strict privacy guarantees.

However, for many real-world location-based applications, especially continuous ones, imprecise locations via obfuscation are not feasible due to the high accuracy requirements of location data. For such location-based applications which need real locations, anonymizing locations by removing/changing real IDs to prevent trajectory inference is a viable approach [15], [16], [17], [18], [19], [20]. Various anonymity methods have been proposed to protect human mobility trajectories, ranging from trajectory anonymity (using a fixed pseudonym for the whole trajectory) [16], [18] to location anonymity (using different pseudonyms for each location) [21], [22], [23]. The location anonymity seems to be a robust privacy-preserving technology, where an entire trajectory is segmented into distinct locations, each assigned a unique pseudonym.

In recent years, concerns have grown that trajectories can still be reconstructed even if location anonymity is deployed. To comprehend this risk, several attack algorithms have been proposed [23], [24], [25]. Some studies even claim nearly 100% recovery accuracy, raising doubts about the effectiveness of location anonymity. However, these attack algorithms have only been validated on simulated datasets generated by simulated tools (*e.g.*, SUMO). **Due to the absence of evaluations on real-world systems, it remain unanswered that whether these simulation results still hold in real world.**

In this paper, we conduct the first real-world study on the effectiveness of location anonymity. One key observation is that emerging online takeout platforms offer a suitable environment for such real-world evaluations. Typically, a takeout service ecosystem consists of multiple interest groups, namely *Consumer*, *Merchant*, *Rider*, and *Takeout Platform*.

As depicted in Fig. 1, mainstream takeout platforms, such as Meituan, Ele.me, Uber Eats, and Gojek, incorporate a feature that displays the real-time locations of nearby riders, known as the **nearby riders feature**. Through reverse engineering of the nearby riders APIs, we discovered that two largest platforms, Meituan and Ele.me, have adopted location anonymity to safeguard riders' trajectories. For the consumer-specified location, these two platforms display the number of nearby riders and their real-time GPS coordinates, excluding their real IDs. It is akin to changing pseudonyms every time a rider's location is displayed, *i.e.*, using location anonymity.

We firstly design a large-scale data collection system to gather anonymized location data of riders from Ele.me and Meituan for real-world evaluations. Achieving this is challenging, as these platforms conceal the details of nearby riders API, implement anti-collection measures, and restrict the amount of data collected by single collector. To overcome these challenges, we designed a large-scale data collection system with modules of *Nearby Riders API Pinpointing and Parsing*, *Anti-collection Mechanisms Bypassing*, and *Parallel Collection*. Using this system, we collected anonymized location data of riders from Ele.me and Meituan within a $4.5km$ x $9km$ area, lasting for 14 days for ele.me and 8 days for meituan, forming two anonymized datasets containing millions of riders' locations. Additionally, to collect labelled datasets (ground-truth datasets), we randomly recruited 20 Ele.me and 8 Meituan riders in our experimental area, recording their trajectories. The anonymized datasets and ground-truth datasets present an occasion for comprehensively evaluating the effectiveness of location anonymity in practice.

Our preliminary analysis reveals that some key factors, such as indoor positioning deviation and movement randomness of riders within gathering areas (*e.g.*, shopping centers), make massive riders' trajectories appear messy in these areas (as shown in Fig. 2). This leads to existing trajectory inference algorithms, which only rely on the correlation of locations in adjacent time slots, fail to accomplish the inference task for massive riders with messy trajectories. To address this, we propose an innovative multi-stage trajectory inference attack specifically tailored to location anonymity. We treat the gathering area as a blackbox where locations are ignored and divide the trajectory inference task into multiple stages. First, we propose a location linking algorithm that associates locations belonging to the same rider by constructing reasonable transition probability matrix. The algorithm stops linking if a rider enters the gathering area, generating multiple segments of a rider's trajectory instead of the full trajectory. Then, we further design a segment matching algorithm to identify segments belongs to the same rider through feature comparison of segments, enabling cross-gathering area tracking.

We initially assess the performance of location anonymity using static metrics (*e.g.*, mix set), followed by evaluations conducted through the proposed multi-stage trajectory inference attack, to form the comprehensive evaluations. We give the concrete definition of the mix set in the context of location anonymity, and calculate the mix ability under various situations. Experimental results indicate that 89.3% of riders' locations have at least one mix neighbor, and even 15.6% locations have at least nine mix neighbors. Then we implement the proposed multi-stage trajectory inference attack on the anonymized datasets and conduct extensive evaluations. The results refute the effectiveness of location anonymity for short-term tracking (achieving 83.4% and 74.5% inference accuracy on Ele.me and Meituan) but confirm its utility for long-term tracking. Experimental analysis reveals that strong aggregation properties of riders formed by the characteristics of takeout services (riders need to frequently enter and exit gathering areas for picking up goods), previously deemed unrealistic across multiple scenarios, play a crucial role in thwarting long-term tracking, providing valuable insights for location privacy protection.

The contributions of our study include:

- **Large-scale data collection system**: We design a large-scale data collection system to gather anonymized locations of riders from Ele.me and Meituan, resulting in two anonymized datasets comprising millions of riders' locations.
- **Multi-stage trajectory inference attack**: We propose a multi-stage trajectory inference attack, which incorporates a location linking stage that associates locations belonging to the same rider to form multiple segments outside the gathering area, as well as a segment matching stage to find segments belongs to the same rider.
- **Measurement and Evaluation**: We assess the performance of location anonymity using static metrics and the proposed multi-stage trajectory inference attack, to form a comprehensive evaluation. Extensive evaluations refute the effectiveness of location anonymity for short-term tracking but confirm its utility for long-term tracking.

The remainder of this paper is organized as follows. Sec. II introduces the background. Sec. III describes the large-scale data collection system. Sec. IV and Sec. V present the proposed attack and evaluation results. We discuss the related work and conclude this paper in Sec. VI and Sec. VII.

II. Overview of Takeout Services

The takeout service market has experienced substantial growth in the past five years. According to the statistics, China leads with the largest user base and market penetration. Driven by huge economic benefits in takeout service market, many takeout platforms have been founded, including Uber Eats, Ele.me, and Meituan.

Nearby riders feature. A takeout service ecosystem typically includes multiple interest groups, including *Consumer*, *Merchant*, *Rider*, and *Takeout Platform*. Delivery riders are an important part of the online to offline pattern. A rider will jointly consider the locations of all users (merchants and consumers) to choose the shortest delivery path. To demonstrate the service capabilities, mainstream takeout platforms incorporate a nearby riders feature on their consumer apps that displays the real-time locations of nearby riders.

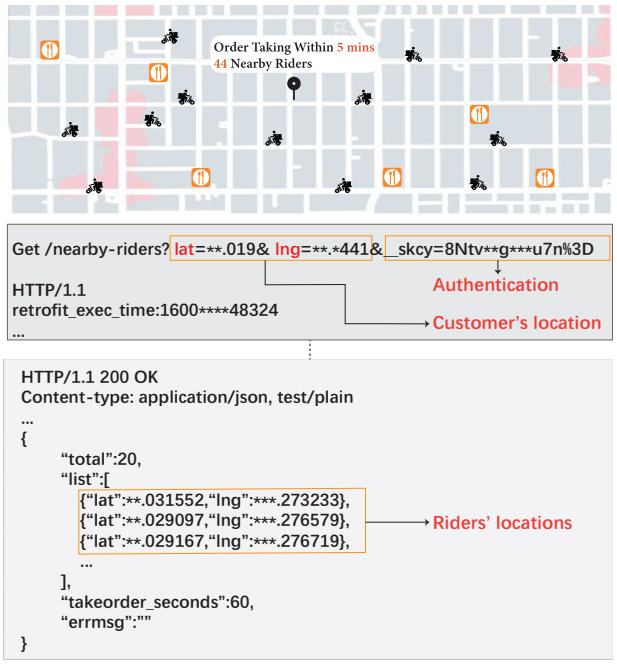


Fig. 1: An example of nearby riders display

Takeout platforms vary in displaying nearby riders due to regional population density and user preferences. In Europe, the display of nearby riders is restricted due to strict privacy regulations. In the United States, some platforms (*e.g.*, Uber Eats) offer a brief history path of riders. In Asia, two largest takeout platforms in the world (*i.e.*, Meituan and Ele.me) reveal the real-time GPS coordinates of nearby riders.

Fig. 1 illustrates an example of the nearby riders feature available on Ele.me and Meituan. When a consumer specifies a location on the consumer app, the information of nearby delivery riders who could provide services to the consumer within a fixed time will be displayed, including the total number of nearby riders and their GPS coordinates without real IDs.

III. Large-scale Data Collection System

In this section, we outline the method for large-scale data collection using Nearby Riders APIs and present the datasets collected from two takeout platforms (*i.e.*, Ele.me and Meituan), which could be used for the real-world evaluations of location anonymity. It is not trivial to achieve large-scale data collection from these platforms due to following challenges:

- **Undocumented APIs.** Contrary to system APIs that provide detailed explanations, takeout platforms do not disclose any information about the specifics of the Nearby Riders APIs, which increases the difficulty of actively performing Nearby Riders API calls to obtain data from designated locations.
- **Anti-collection Mechanisms.** Like other service providers, takeout platforms also implemented some anti-collection mechanisms on the Nearby Riders APIs to prevent attackers from disrupting the operation of the platforms.
- **Selective Responses.** Platforms also impose restrictions on the amount of data that a single collector can obtain, such as setting maximum search radius and limiting the number of displayed nearby riders.

A. System Design

To solve the above challenges, we propose a large-scale data collection system consisting of three modules: the Nearby Riders API Pinpointing and Parsing, Anti-collection Mechanisms Bypassing, and Parallel Collection.

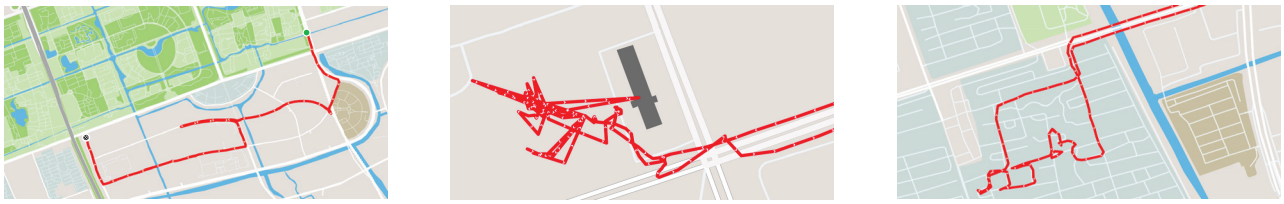
1) The Nearby Riders API Pinpointing and Parsing: To solve the first challenge (*i.e.*, undocumented APIs), we pinpoint Nearby Riders APIs from massive request and response messages. Firstly, we change the location on the consumer app to trigger Nearby Riders API calls, and use an analysis tool, "Packet Capture", to capture all messages between the consumer app and takeout platform. Then we search for related messages of Nearby Riders API based on the following observations: the request message contains consumer-specified location, and the response message of Nearby Riders API contains a series of GPS coordinates. For the convenience of pinpointing, we halt all other actions on the app except for changing locations.

Through parsing the related messages of Nearby Riders API based on special characters (*e.g.*, &), we discovered that a request message mainly contains the request URL, consumer' location, and authentication field, while a response message mainly contains the number of displayed nearby riders, GPS coordinates of riders, and estimated time for order pickup.

2) Anti-collection Mechanisms Bypassing: After multiple attempts, we have identified some anti-collection mechanisms used to block or slow down requests from non-consumer groups. We will discuss methods for bypassing these mechanisms to execute Nearby Riders API calls with spoofed GPS coordinates.

Bypassing Authentication. The first mechanism employed by takeout platforms is authentication, which restricts access to the Nearby Riders APIs to authenticated consumers only. Through in-depth analysis of the authentication field in request messages, we have identified two types of authentication mechanisms adopted by takeout platforms:

- *Location-independent authentication.* Ele.me adopts the location-independent authentication. A cookie will be served as a mean of authentication for subsequent HTTP requests (including Nearby Riders API). Since cookie-based authentication is independent of current status of consumers (*e.g.*, consumers' locations), attackers could obtain valid cookies to continuously perform Nearby Riders API calls with spoofed GPS coordinates. Notably, cookies did not expire when *userid* is set to 0, significantly reducing the difficulty for attackers.
- *Location-dependent authentication.* Meituan employs location-dependent authentication, generating unique



Rider's delivery route on half way

Rider's delivery route passing through shopping center

Rider's delivery route passing through neighbourhood

Fig. 2: A rider's route passing through different areas

signatures for each location and adding them to the authentication filed, to check the consistency of the locations in the signatures and the consumer-specified locations. Through decompiling Meituan APK, we discovered that a signature can be generated by passing the consumer's location, client_key, and UUID to the *get_signature* function. Since the *libmtguard.so* file called by this function uses some unknown parameters which are protected by code obfuscation, it is difficult to forge valid signatures for spoofed locations. Despite Meituan adopts a stronger authentication mechanism, we have still identified overlooked vulnerabilities that can be used to bypass it. Since the consistency checking mechanism does not verify timestamps in signatures, we can obtain valid signatures for designated locations by moving a consumer's location to those locations and parsing captured messages. These valid signatures can then be reused for subsequent data collection.

Bypassing Rate Limiting. Rate limiting is commonly used as a countermeasure against DoS attacks [3]. Based on our testing, platforms implements rate limiting by recording the number of requests from each IP and blocking requests when the limit is reached. To perform multiple Nearby Riders API calls simultaneously, we use dynamic IP addresses to avoid triggering rate limiting.

3) Parallel Collection: Using above steps, an attacker can obtain nearby riders' locations of designated locations. However, due to the following limitations, we needs to set multiple collectors at different locations and simultaneously perform API calls for achieving data collection in a large area.

- **Displaying Locations of Partial Nearby Riders:** To prevent attackers from abusing the near riders feature, these takeout platforms limit the number of displayed GPS coordinates. Upon reviewing the response messages, we discover that Ele.me platform shows the actual number of nearby riders around the consumer-specified location, but displays GPS coordinates of no more than 20 of these nearby riders, while Meituan platform limits the display of GPS coordinates to 15 riders.
- **Setting Maximum Search Radius:** Takeout platforms also set a maximum search radius for Nearby Riders API calls. To verify the maximum search radius, we performed Nearby Riders API calls with spoofed GPS coordinates (*i.e.*, consumer-specified locations) and stored locations of nearby riders of a spoofed GPS coordinate as a nearby riders record. The data collection process lasted for one day. Then we selected records that did not trigger the limit on the number of nearby riders, and computed the maximum distance between consumer-specified locations and the GPS coordinates of their nearby riders. Experimental results show that maximum search radius is $3km$ for Ele.me and $5km$ for Meituan.
- **Displaying Locations of Nearby Riders Randomly:** Takeout platforms randomly select nearby riders for display when the number of nearby riders exceeds the limit, rather than following a certain rule (*e.g.*, shortest distance). We designed two experiments to verify this: one consumer sends two API requests with the same location simultaneously, and two consumers send one identical request simultaneously. We conducted experiments on 30 GPS coordinates, lasting for ten minutes. Through analyzing nearby riders records at the same time and location, we found that they contain different GPS coordinates. For collected data from Ele.me, the similarities of nearby riders records at the same time and location are 67.70% and 37.39% in two experiments. For Meituan, because the format of request packets are not applicable to different consumers, we only conducted the first experiment, indicating a similarity of only 26.75%.

Due to the above mechanisms, we need to deploy multiple collectors to gather riders' data in a large area, by simultaneously and continuously performing API calls. Since takeout platforms set the maximum search radius, a straightforward way of deploying collectors is to divide the large area into many sub-grids with an edge length of $2 \times$ *maximum radius*, and place collectors in the center of each sub-grid. However, takeout platforms also limit the number of displayed GPS coordinates of nearby riders and randomly display nearby riders, which may lead to missing riders' location data if using this edge length. To collect more data, redundant collectors are necessary. To obtain the appropriate edge length, we varied the edge length from $100m$ to $1,000m$ and collected data for one day in a $5km \times 5km$ area. Experimental results show that we could collect most data with an edge length of $300m$ (Ele.me) and $150m$ (Meituan).

B. Data Corpus

To evaluate the performance of the location anonymity method in practice, we conducted long-term data collection to form the large-scale and high quality datasets.

Data collection system deployment. Based on survey information from some riders, the typical delivery range (or activity range) of riders is centered at a rider-specified location, usually a shopping mall. Therefore, we chose an area with dimensions of $4.5km \times 9km$, centered around a prosperous shopping mall, for large-scale data collection, which is sufficient to cover the trajectories of the riders surrounding the shopping mall.

Using the methodology outlined in Sec. III-A, we divided the large area into many sub-grids and positioned collectors at the center of each grid. Once deployed, these collectors began gathering data at a consistent request rate. For Ele.me, we discovered that cookies do not expire when *userid* is set to 0, allowing us to use a valid cookie for continuously performing API calls with spoofed GPS coordinates. In the case of Meituan, we need to collect signatures of designated locations of collectors in advance.

Anonymized Datasets: We collected anonymized location data of riders from Ele.me and Meituan over 14 days and 8 days, respectively. Due to duplication of some data collected by adjacent collectors at the same time, we removed these duplicates before aggregating the data into the format: <timestamp, GPS coordinates set>. This process resulted in two anonymized datasets, containing millions of riders' locations.

Ground-truth Datasets: Meanwhile, to verify the correctness of our trajectory inference algorithms, we recruited 20 riders from Ele.me and 8 riders from Meituan within our experimental area. Each volunteer carried a GPS-tracking device to record their trajectories. We have obtained authorization from our university for these experiments.

IV. Multi-stage Trajectory Inference Attack

In this section, we define the threat model to describe the ability of the adversary, and design a multi-stage inference attack for further evaluating the performance of the location privacy protection achieved by location anonymity.

A. Threat model

In our attack framework, the adversary seeks to reconstruct riders' trajectories from anonymized locations, which can be regarded as a multiple-target tracking problem. Specifically, a variable number of riders move within a given area, and their locations are displayed at irregular intervals. A snapshot of multiple riders' locations at time slot t is defined as a set $L^t = \{l_1^t, l_2^t, l_3^t, ...\}$, where l_i^t denotes a rider's location at time slot t. The trajectory inference algorithm associates these locations with subsequent anonymized locations to form appropriate trajectories, *i.e.*, sequences of locations belong to the same rider.

As depicted in Fig. 2, a rider's trajectory is messy in the gathering area (*e.g.*, shopping centers), which is caused by the following reasons: 1) GPS accuracy diminishes within buildings, and 2) riders often move irregularly to pick up multiple orders. Thus, it is difficult to recover a rider's trajectory from massive riders with messy trajectories in the gathering area. Therefore, we consider a multi-stage trajectory inference attack by regarding gathering area as a blackbox where locations are ignored, and dividing the inference process into multiple stages.

In the first stage, the adversary aims to link locations in L^t with locations in subsequent anonymized location sets to form appropriate sub-trajectories, composed of locations belonging to a rider. The linking process stops if a rider enters the gathering area. With these steps, the adversary could recover riders' sub-trajectories (*i.e.*, segments) outside gathering areas. In the second stage, the adversary will link sub-trajectories of entering and exiting gathering area. Let $Tra_{in} = \{tra_{in_1}^{t_1}, tra_{in_2}^{t_2}, \cdots\}$ be sub-trajectories of entering a gathering area and $Tra_{out} = \{tra_{out_1}^{k_1}, tra_{out_2}^{k_2}, \cdots\}$ be sub-trajectories of exiting a gathering area, where t_i and k_j represents entering and exiting times. In this stage, the adversary wants to find sub-trajectory pairs of the same rider, including an entering sub-trajectory and an exiting sub-trajectory.

B. Attack Framework

Given the challenge of reconstructing a rider's trajectory from massive riders with disorganized trajectories in the gathering areas (such as shopping malls), we propose an innovative multi-stage inference attack specifically tailored to address these challenges, as shown in Fig. 3. Firstly, we propose a location linking algorithm that associates locations belonging to the same rider by constructing reasonable transition probability matrix, and stops location linking if a rider enters the gathering area. Due to ignoring rider's locations in the gathering areas, we could only generate multiple sub-trajectories (segments) for riders instead of the entire trajectory of them. To address this, we further design a segment matching algorithm to identify sub-trajectories belongs to a same rider by feature comparison.

1) Location Linking Algorithm for Non-gathering areas: In this stage, the adversary focuses solely on locations outside gathering areas. Firstly, the adversary link the locations in L^t with those in subsequent anonymized location sets $L^{t+\{1,...,\Delta t\}} = \{L^{t+1}, \ldots, L^{t+\Delta t}\}$, which can be modeled as constructing a weighted bipartite graph. Each edge is weighted based on the probability that a location l_i^t and a subsequent location $l_j^{t+\{1,...,\Delta t\}} \in L^{t+\{1,...,\Delta t\}}$ belong to the same rider. We consider the Δt subsequent anonymity location sets simultaneously to accommodate situations where riders may disappear for some time slots. Therefore, the adversary needs to calculate the probability matrix $A^{t \to t+\{1,...,\Delta t\}} = \{a_{ij}^{t \to t+\{1,...,\Delta t\}}\}_{m \times n}$, where $a_{ij}^{t \to t+\{1,...,\Delta t\}}$ represents the the probability that a location l_i^t and a subsequent location $l_j^{t+\{1,...,\Delta t\}}$ belong to the same rider. Consequently, this problem is transformed into how to accurately build the transition probability matrix $A^{t \to t+\{1,...,\Delta t\}}$.

The continuity of human mobility could make us estimate the next location by using the current location [26], [27]. We build the mobility model as follows:

$$d_i^{t+k} = \|l_i^t - l_i^{t-1}\| \times k, k \in \{1, 2, \ldots, \Delta t\} \quad (1)$$

where d_i^{t+k} represents the theoretical distance that rider i moves from the time slot t to the time slot $t+k$, and $\|.\|$ represents the distance traveled within a time slot, which is computed by historical average speed. Based on the theoretical moving distance, we can calculate the subsequent theoretical locations for each time slot. For each subsequent location l_j^{t+k}, where $k \in \{1, \ldots, \Delta t\}$, if the distance between l_j^{t+k} and l_i^t exceeds $\Delta d_{max} \times k$ (Δd_{max} is the maximum movement distance within a time slot), the transition probability $a_{ij}^{t \to t+k}$ will be set to zero. For other locations that satisfy the maximum distance limit, we can calculate the distance between them and the subsequent theoretical locations, and set the transition probability $a_{ij}^{t \to t+k}$ based on the following rule: The smaller the distance, the higher the transition probability.

When analyzing the ground-truth data of riders, we observed that the probability of the same location appearing within adjacent time slots belongs to different riders is very low. Therefore, if a location identical to l_i^t exists in $L^{t+\{1,\ldots,\Delta t\}}$, we assign a transition probability of one to it and set the rest to zero. For each location $l_j^{t+\{1,\ldots,\Delta t\}}$ where $a_{ij}^{t \to t+\{1,\ldots,\Delta t\}} \neq 0$, we record the linkage between the location $l_j^{t+\{1,\ldots,\Delta t\}}$ and l_i^t, and continually execute the location linking algorithm by setting the location $l_j^{t+\{1,\ldots,\Delta t\}}$ as the current location. If a location $l_j^{t+\{1,\ldots,\Delta t\}}$ is not selected as the subsequent location of a trajectory, it will be regarded as the starting location of a new trajectory.

By following the aforementioned steps, the adversary could acquire numerous segments (*i.e.*, sub-trajectories). Since we record all possible location linkages, some of these segments may share the same starting location. Therefore, for all segments having the same starting location, we calculate the segment probabilities based on the transition probabilities of location linkage within a segment and only retain the segment with the highest probability.

By utilizing the location linking algorithm, we can generate a substantial number of trajectories. To minimize computational costs, we analyzed both the recovered trajectories and the actual trajectories of the riders in the ground-truth datasets, and summarized their characteristics. This analysis aids in excluding improbable trajectories. We observed that riders in real-world scenarios always exhibit the following characteristic in their trajectories: Riders maintain a discernible overall direction of travel during the delivery process. Except for the designated pickup locations and order addresses, they do not exhibit back-and-forth movements along the trajectory. Based on the above observation, we can eliminate unreasonable trajectories when inferring, significantly reducing the overhead of the algorithm.

2) Segment Matching Algorithm for gathering areas: With the above steps, the adversary could recover riders' sub-trajectories between gathering areas. In the second stage, the adversary will link the sub-trajectories of entering and exiting the gathering areas.

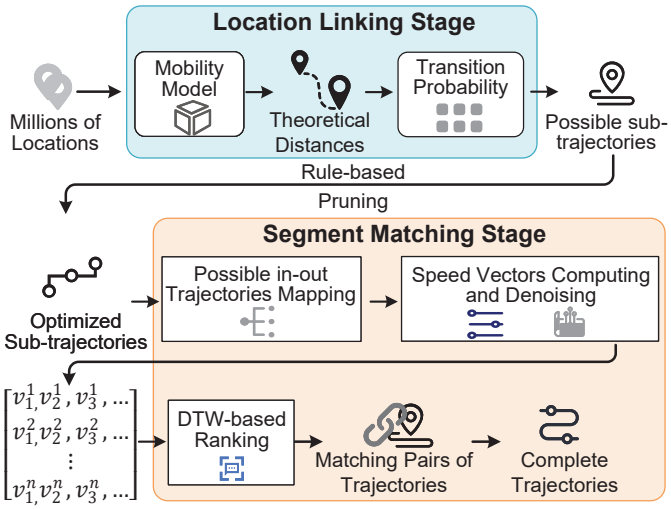


Fig. 3: Attack Framework

Previous works have pointed that two sub-trajectories belonging to the same person often exhibit similar patterns [28], such as similar speed change trends and the higher probability of visiting same locations. Due to the nature of riders' locations being determined by consumers' orders, sub-trajectories of a rider seldom share the same visited locations. Fortunately, upon observing 50 real pairs of entering and exiting sub-trajectories in the ground-truth datasets, we discovered that the speed values between entering and exiting sub-trajectories of the same rider exhibit similar features. This phenomenon is due to following reasons: 1) Riding habits of riders. Each rider possesses unique riding habits, which are reflected in their speed values. For the same rider, there will be similar riding habits (*i.e.*, speed patterns) in the entering and exiting trajectories. 2) Battery level of electromobile: electromobile's speed is highly dependent on its current battery level. Consequently, when a rider enters and exits a gathering area, they tend to maintain a relatively stable speed pattern. Furthermore, differences in battery levels among different riders result in distinct speed patterns, which can be utilized to distinguish between riders.

Based on the above insights, we propose a segment matching algorithm that relies on feature comparison. The fundamental idea is to construct the speed vector for each entering/exiting sub-trajectory and compute the speed similarities for all pairs of entering and exiting sub-trajectories. For an entering sub-trajectory, we could determine all possible exiting sub-trajectories based on temporal correlation (the interval between enter and exit time is less than 20 minutes) and spatial correlation (belonging to the same gathering area). For any sub-trajectory, we can obtain a series of speed values between two locations, and generate the speed vector $V = \{v_1, v_2, v_3, \cdots\}$. For each speed vector V of a sub-trajectory,

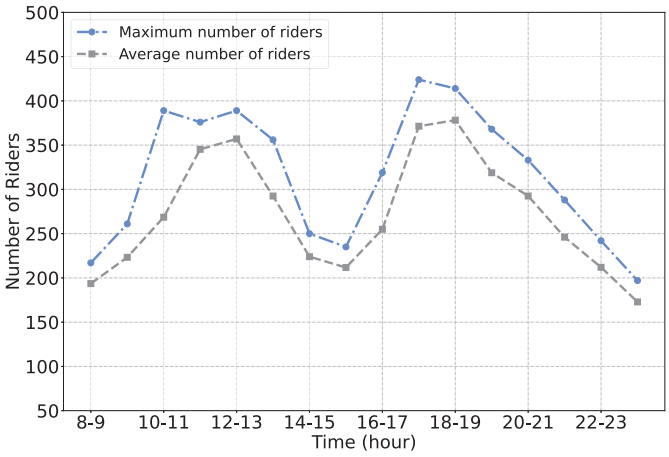


Fig. 4: Number of riders under different times

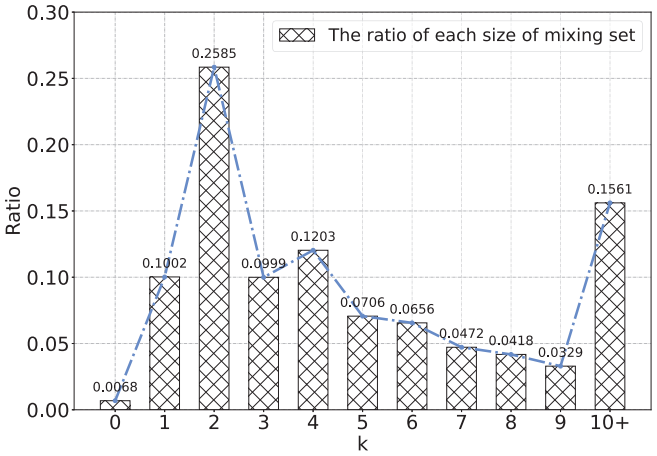


Fig. 5: Ratio of riders' locations under each size of mix set

we need to remove noises caused by traffic emergency, such as the sudden appearance of a pedestrian, which could by solved by elliptic envelope method. To compute the speed similarities (*i.e.*, the segment matching probability) between a pair of entering and exiting sub-trajectories, we adopt the Dynamic Time Warping (DTW) algorithm. DTW is a method that calculates the difference between two given sequences of varying lengths. For each entering sub-trajectory, we compute the DTW distance between its speed vector and the speed vectors of all possible exiting sub-trajectories. We then rank the exiting sub-trajectories based on the DTW distance: The smaller the distance, the higher the likelihood that it is the exiting sub-trajectory corresponding to the entering sub-trajectory. Based on the ranking, we can link sub-trajectories that span one or more gathering areas to form the complete trajectory of riders.

V. Measurement and Evaluation

In this section, we firstly use static metrics to evaluate location anonymity on the real-world datasets, followed by the evaluations using the proposed multi-stage trajectory inference attack to fully assess the effectiveness of the location anonymity in practice.

A. Static Measurement of Location Anonymity

The mix set is an useful metric for assessing the extent of location privacy protection achieved by location anonymity strategies. In the context of location anonymity, we define the mix set as follows: For a rider u_i, let $l^t_{u_i}$ represent the location of u_i at time slot t. Take this location as the center and the maximum movement distance Δd_{max} as the radius, we can determine the maximum movement range of u_i in the next time slot. All riders within this circular area at next time slot could potentially be the rider u_i, forming the mix set of the rider u_i. We use the mix set as the metric for evaluation on a day's data from the Ele.me platform. To save space, the results on Meituan platform are not shown, as these two platforms exhibit similar situations.

As explained in Sec. III-A, whether the location of a rider is displayed within a time slot is random, leading to variability in collected locations of riders across time slots. Therefore, we firstly calculated the number of riders under different times to represent the distribution of riders in our datasets, as illustrated in Fig. 4. Experimental results show that there are over 150 active riders (average) in each time slot within our collection area, and the number can even reach up to 300 at lunch/dinner times, proving that it can serve as a high-quality dataset for evaluating location anonymity.

Then we calculated the overall mix ability, *i.e.*, the size of mix set of each location. Fig. 5 shows the ratio of riders' locations under each size of mix set. 89.3% riders' locations have at least one mix neighbor, and even 15.6% locations have at least nine mix neighbors. We also evaluated the size of mix set from temporal dimension, as shown in Fig. 6, which displays the distribution of mix set size across different times. We could find that large mix sets (*e.g.*, ≥ 6) always appear during lunch/dinner times. This phenomenon is related to the characteristics of platform services and people's living habits, as a large number of riders frequently enter and exit gathering areas to pick up goods and move along roads for delivery during lunch/dinner times.

City level comparison. We also conducted experiments using daily data of three cities to demonstrate there is no contingency in the above experimental results. Using the comparable experimental environments (*e.g.*, the size of the experimental area) and the designed data collection system, we also collected anonymized location data of riders from other two cites. Then we processed these location data of riders and computed the mix sets of them. Fig. 7 shows the distribution of mix sets across three cities. While there are fluctuations in the ratio of different sizes of mix sets across different cities due to variations in population density and lifestyle habits, location anonymity still demonstrates consistent and effective protection in different cities.

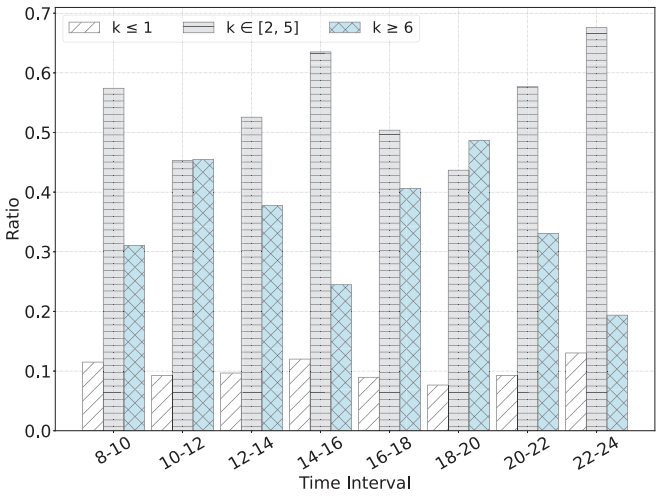


Fig. 6: The distribution of mix set size under different times

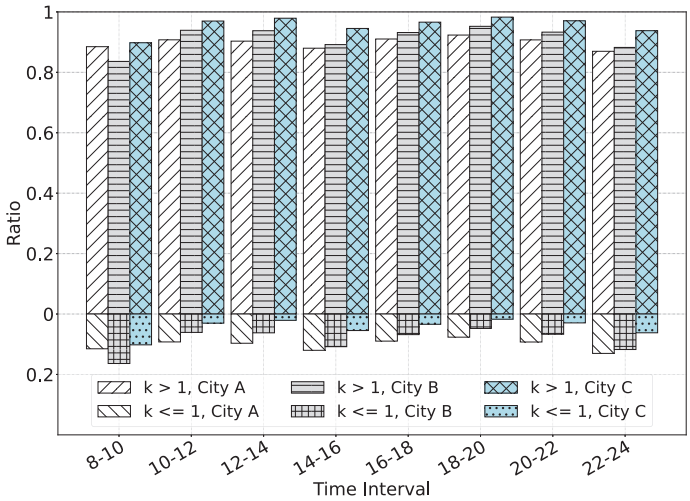


Fig. 7: The distribution of mix sets across different cities

B. Dynamic Evaluations through Multi-stage Trajectory Inference Attack

In this section, we systematically evaluated the performance of location anonymity through the proposed attack. We applied our attack on anonymized datasets (as described in Sec. III-B), which comprised millions of anonymized locations of riders.

1) Metrics: Similar to [28], we use two well-known metrics, time-traceability and distance-traceability, for performance evaluations. Traceability metrics can quantify the degree to which a user can be traced over time or distance. The inferred trajectories from anonymized datasets and real trajectories of riders (ground-truth datasets) are represented by discrete locations. Due to different collection frequencies and GPS deviation, the same road segment in these datasets may be represented by different locations. For the convenience of comparison, we transformed the representation of discrete locations into the representation of roads using APIs of navigation systems. Let $\widetilde{Tra_i} = (\widetilde{road_0}, \widetilde{road_1}, \widetilde{road_2}, \cdots)$ denote inferred road segments, and $Tra_i = (road_0, road_1, road_2, \cdots)$ denote real road segments of the rider i. The time- and distance-traceability metrics are defined as follows:

$$p_t(Tra_i, \widetilde{Tra_i}) = \frac{\sum_{k=0}^{n} t_{used}(road_k) \cdot \mathcal{I}(road_k)}{t_{used}(Tra_i)} \quad (2)$$

$$p_d(Tra_i, \widetilde{Tra_i}) = \frac{\sum_{k=0}^{n} d_{traveled}(road_k) \cdot \mathcal{I}(road_k)}{d_{traveled}(Tra_i)} \quad (3)$$

Here, t_{used} and $d_{traveled}$ represent the used time and traveled distance of a road segment/trajectory. $\mathcal{I}(road_k) = 1$ if $road_k$ can be correctly recovered. Otherwise, $\mathcal{I}(road_k) = 0$.

2) The performance of Location Linking Algorithm for Non-gathering areas: Using time- and distance-traceability metrics, we evaluated the performance of location linking algorithm on two anonymized datasets. The results are shown in Table I. Experimental results show that location linking algorithm achieves a success rate of over 83.4% with the distance-traceability metric and around 72.1% with the time-traceability metric in tracking riders on Ele.me platform. On the Meituan platform, the success rate is over 74.5% with the distance-traceability metric and around 66.4% with the time-traceability metric. Our evaluation demonstrates that the adversary could achieve relatively high accuracy on both platforms when tracking riders in non-gathering areas. The success rate on the Meituan platform is slightly lower due to the fact that the platform does not provide services in some sub-regions of our collection area, resulting in uncollectable location data for a portion of riders.

TABLE I: The performance of location linking algorithm

Platforms	Non-gathering areas	
	time-traceability	distance-traceability
Ele.me	72.1%	83.4%
Meituan	66.4%	74.5%

Although we cannot achieve long-term tracking through this step (location linking algorithm), the inferred sub-trajectories can still lead to privacy leakage, *e.g.*, inferring rider's homes by aggregating starting routes of multiple days and deducing order information (which residential zone's consumers placed orders from which merchants or shopping malls) from the sub-trajectories. The sizes of mix sets in non-gathering areas are significantly lower than the sizes of mix sets in gathering areas. Insufficient mix level is the primary cause of location privacy leakage. Without sufficient mix locations, some side channel information, such as driving direction and road information, will reveal the relationship between two anonymized locations. Due to the high mobility of riders on the roads, it is difficult to form uniform and large-scale mix locations in non-gathering areas, making location anonymity susceptible to the influence of side channel information and less ineffective.

3) The Performance of Segment Matching Algorithm in gathering areas: In second stage, we links the sub-trajectories of entering and exiting the gathering areas using the segment

matching algorithm. We evaluated the performance of segment matching algorithm on the results of the previous step, *i.e.*, inferred sub-trajectories of entering and exiting gathering areas. For the sub-trajectories, we can generate the speed vector and use DTW algorithm to compute the speed similarities (*i.e.*, the segment matching probability).

Due to the influence of unstable speed and the ever-changing road conditions, it is challenging to accurately select the real exiting sub-trajectory in practice. Therefore, we used a new metric, top-k trajectories, to evaluate the matchability. Experimental results on the Ele.me platform show that when selecting top-10 candidate exiting sub-trajectories, there exist a trajectory that can cover 77.8% of the real trajectory. For the Meituan platform, the value is 68.9%. When a rider enters and exits a gathering area, we still have a certain probability of inferring the rider's trajectory. However, when a rider across multiple gathering areas, the inference errors rapidly increase, causing the inability to long-term tracking.

C. Takeaway

In our experiments, we used static metrics and dynamic attacks to evaluate the performance of location anonymity on the real-world datasets. Static measurement results show that location anonymity appears to be performing well, as 89.3% riders' locations have at least one mix neighbor, and even 15.6% locations have at least nine mix neighbors. Unfortunately, location anonymity is susceptible to the influence of side channel information. Some side channel information, such as driving direction and road information, will leak the relation between two anonymized locations, resulting in a higher success rate of short-term tracking. The emergence of gathering areas (*e.g.*, shopping malls) where massive riders exhibit messy trajectories helps to disrupt the long-term tracking of riders' trajectories. When a rider across multiple gathering areas, trajectory tracking has become almost impossible. This strong aggregation properties of riders formed by the characteristics of takeout services (*i.e.*, frequently entering and exiting multiple gathering areas for picking up goods), previously deemed unrealistic across multiple scenarios, play a crucial role in thwarting long-term tracking, offering valuable insights for designing more effective location privacy protection solutions.

VI. Related Work

Location privacy has received significant attention in recent years due to the widespread adoption of location-based services [29], [30], [31], [32]. Researchers have investigated various approaches to safeguard location privacy, such as location generalization, location perturbation, and location anonymization.

Location generalization refers to hiding the location in a way that decreases its precision or detail, making it less precise while still retaining a certain level of usability [33], [34], [35]. Location perturbation is a crucial strategy in the field of location privacy protection, which can prevent an adversary from inferring the exact location by introducing random noise. In recent years, differential privacy (DP), a highly rigorous privacy model, has been widely used to protect location privacy. Andrés *et al.* proposed geo-indistinguishability [11] and inspired a lot of subsequent works [12], [13], [36]. Through years of development, these approaches have evolved from manually setting the simple rules (*e.g.*, offsetting a specific distance) to adding random noises that comply with strict privacy guarantees.

Another line of research in the field of location privacy focuses on anonymizing locations/trajectories by removing/changing real IDs. Various anonymity techniques have been suggested to safeguard human mobility trajectories. Initially, trajectory anonymity techniques that use a fixed pseudonym for the whole trajectory were employed by several applications. However, some works have demonstrated that users can still be re-identified using external information [15], [16], [18]. The mix-zone is a classic model designed to preserve location privacy in services that involve continuous location exposure. It effectively breaks the continuity of a user's location exposure by frequently changing pseudonyms within designated areas [17], [19], [20], [28]. Further, location anonymity, which uses different pseudonyms for each location, strictly divides an entire trajectory into distinct locations with different pseudonyms [21], [22], [23]. However, current works only verified the effectiveness of location anonymity on the simulated datasets generated by simulated tools, which may differ significantly from the actual scenarios. The effectiveness of location anonymity in practical scenarios remains uncertain and has received less attention so far.

VII. Conclusion

In this paper, we investigate the effectiveness of location anonymity in practical scenarios (*i.e.*, two popular takeout platforms). We firstly design a large-scale data collection system and collect millions of anonymized location data of riders from two takeout platforms. We then propose an innovative multi-stage trajectory inference framework specifically tailored to study the effectiveness of location anonymity. The experiments conducted on the collected datasets refute the effectiveness of location anonymity for short-term tracking but confirm its utility for long-term tracking. Experimental analysis reveals that strong aggregation properties of riders, previously deemed unrealistic across multiple scenarios, play a crucial role in thwarting long-term tracking.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant 62302362, 62132013, and 62072359, the Key Research and Development Programs of Shaanxi under Grant 2024GX-YBXM-071, the Basic Strengthening Plan Program under Grant 2023-JCJQ-JJ-0772, Key Research and Development Projects of Jilin Province under Grant No. 20240302090GX, and the Fundamental Research Funds for the Central Universities, JLU. Lichuan Ma is the corresponding author (e-mail: lcma@xidian.edu.cn).

REFERENCES

[1] I. Polakis, G. Argyros, T. Petsios, S. Sivakorn, and A. D. Keromytis, "Where's wally? precise user discovery attacks in location proximity services," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 817–828.

[2] W. U. Hassan, S. Hussain, and A. Bates, "Analysis of privacy protections in fitness tracking social networks -or- you can run, but can you hide?" in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, Aug. 2018, pp. 497–512.

[3] Q. Zhao, C. Zuo, G. Pellegrino, and Z. Lin, "Geo-locating drivers: A study of sensitive data leakage in ride-hailing services," in *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*. The Internet Society, 2019.

[4] H. Givehchian, N. Bhaskar, E. R. Herrera, H. R. L. Soto, C. Dameff, D. Bharadia, and A. Schulman, "Evaluating physical-layer ble location tracking attacks on mobile devices," in *2022 IEEE Symposium on Security and Privacy (SP)*, 2022, pp. 1690–1704.

[5] N. Lakshmanan, N. Budhdev, M. S. Kang, M. C. Chan, and J. Han, "A stealthy location identification attack exploiting carrier aggregation in cellular networks," in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Aug. 2021, pp. 3899–3916.

[6] X. Han, J. Xiong, W. Shen, Z. Lu, and Y. Liu, "Location heartbleeding: The rise of wi-fi spoofing attack via geolocation api," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1383–1397.

[7] D. J. Leith and S. Farrell, "Contact tracing app privacy: What data is shared by europe's gaen contact tracing apps," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[8] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2018.

[9] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services: A comprehensive survey," *ACM Comput. Surv.*, vol. 54, no. 1, jan 2021.

[10] M. E. Gursoy, L. Liu, S. Truex, L. Yu, and W. Wei, "Utility-aware synthesis of differentially private and attack-resilient location traces," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 196–211.

[11] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 901–914.

[12] H. Wang, H. Hong, L. Xiong, Z. Qin, and Y. Hong, "L-srr: Local differential privacy for location-based services with staircase randomized response," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2809–2823.

[13] G. Theodorakopoulos, E. Panaousis, K. Liang, and G. Loukas, "On-the-fly privacy for location histograms," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 1, pp. 566–578, 2022.

[14] L. Yu, S. Zhang, L. Zhou, Y. Meng, S. Du, and H. Zhu, "Thwarting longitudinal location exposure attacks in advertising ecosystem via edge computing," in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, 2022, pp. 470–480.

[15] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: Using social network as a side-channel," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 628–637.

[16] H. Wang, C. Gao, Y. Li, G. Wang, D. Jin, and J. Sun, "De-anonymization of mobility trajectories: Dissecting the gaps between theory and practice," in *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, 2018.

[17] B. Palanisamy and L. Liu, "Attack-resilient mix-zones over road networks: architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 14, no. 3, pp. 495–508, 2014.

[18] F. Xu, Y. Li, Z. Tu, S. Chang, and H. Huang, "No more than what i post: Preventing linkage attacks on check-in services," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 620–633, 2021.

[19] I. Bang, T. Kim, H. S. Jang, and D. K. Sung, "An opportunistic power control scheme for mitigating user location tracking attacks in cellular networks," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1131–1144, 2022.

[20] Y. Li, Y. Yin, X. Chen, J. Wan, G. Jia, and K. Sha, "A secure dynamic mix zone pseudonym changing scheme based on traffic context prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9492–9505, 2022.

[21] D. Förster, H. Löhr, A. Grätz, J. Petit, and F. Kargl, "An evaluation of pseudonym changes for vehicular networks in large-scale, realistic traffic scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3400–3405, 2018.

[22] Z. Zhang, T. Feng, B. Sikdar, and W.-C. Wong, "A flickering context-based mix strategy for privacy protection in vanets," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[23] D. Song, R. Tharmarasa, T. Kirubarajan, and X. N. Fernando, "Multi-vehicle tracking with road maps and car-following models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1375–1386, 2018.

[24] J. Wang, Y. Sun, and C. Phillips, "Fake beacon: A pseudonym changing scheme for low vehicle density in vanets," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023, pp. 1–7.

[25] K. Emara, W. Woerndl, and J. Schlichter, "Caps: Context-aware privacy scheme for vanet safety applications," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec '15. New York, NY, USA: Association for Computing Machinery, 2015.

[26] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy in inter-vehicular networks: Why simple pseudonym change is not enough," in *2010 Seventh International Conference on Wireless On-demand Network Systems and Services (WONS)*, 2010, pp. 176–183.

[27] Z. Tu, F. Xu, Y. Li, P. Zhang, and D. Jin, "A new privacy breach: User trajectory recovery from aggregated mobility data," *IEEE/ACM Transactions on Networking*, vol. 26, no. 3, pp. 1446–1459, 2018.

[28] L. Bindschaedler, M. Jadliwala, I. Bilogrevic, I. Aad, P. Ginzboorg, V. Niemi, and J. Hubaux, "Track me if you can: On the effectiveness of context-based identifier changes in deployed mobile networks," in *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012*. The Internet Society, 2012.

[29] E. Bitsikas, T. Schnitzler, C. Pöpper, and A. Ranganathan, "Freaky leaky sms: extracting user locations by analyzing sms timings," in *Proceedings of the 32nd USENIX Conference on Security Symposium*, ser. SEC '23. USA: USENIX Association, 2023.

[30] H. Farrukh, R. Mohamed, A. Nare, A. Bianchi, and Z. B. Celik, "LocIn: Inferring semantic location from spatial maps in mixed reality," in *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 877–894.

[31] E. C. Rye and R. Beverly, "Ipvseeyou: Exploiting leaked identifiers in ipv6 for street-level geolocation," in *2023 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, may 2023, pp. 3129–3145.

[32] K. Gao, H. Wang, H. Lv, and P. Gao, "Your locations may be lies: Selective-prs-spoofing attacks and defence on 5g nr positioning systems," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, 2023, pp. 1–10.

[33] W. Tong, Y. Tong, C. Xia, J. Hua, Q. Li, and S. Zhong, "Understanding location privacy of the point-of-interest aggregate data via practical attacks and defenses," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 3, pp. 2433–2449, 2023.

[34] J. Kang, D. Steiert, D. Lin, and Y. Fu, "Movewithme: Location privacy preservation for smartphone users," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 711–724, 2020.

[35] S. Boukoros, M. Humbert, S. Katzenbeisser, and C. Troncoso, "On (the lack of) location privacy in crowdsourcing applications," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1859–1876.

[36] L. Wang, D. Zhang, D. Yang, B. Y. Lim, X. Han, and X. Ma, "Sparse mobile crowdsensing with differential and distortion location privacy," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2735–2749, 2020.