A Zero Auxiliary Knowledge Membership Inference Attack on Aggregate Location Data

Vincent Guan*
Imperial College London
vincentguan23@gmail.com

Florent Guépin*
Imperial College London
florent.guepin@imperial.ac.uk

Ana-Maria Cretu
EPFL
ana-maria.cretu@epfl.ch

Yves-Alexandre de Montjoye
Imperial College London
demontjoye@imperial.ac.uk

ABSTRACT

Location data is frequently collected from populations and shared in aggregate form to guide policy and decision making. However, the prevalence of aggregated data also raises the privacy concern of membership inference attacks (MIAs). MIAs infer whether an individual's data contributed to the aggregate release. Although effective MIAs have been developed for aggregate location data, these require access to an extensive auxiliary dataset of individual traces over the same locations, which are collected from a similar population. This assumption is often impractical given common privacy practices surrounding location data. To measure the risk of an MIA performed by a realistic adversary, we develop the first Zero Auxiliary Knowledge (ZK) MIA on aggregate location data, which eliminates the need for an auxiliary dataset of real individual traces. Instead, we develop a novel synthetic approach, such that suitable synthetic traces are generated from the released aggregate. We also develop methods to correct for bias and noise, to show that our synthetic-based attack is still applicable when privacy mechanisms are applied prior to release. Using two large-scale location datasets, we demonstrate that our ZK MIA matches the state-of-the-art Knock-Knock (KK) MIA across a wide range of settings, including popular implementations of differential privacy (DP) and suppression of small counts. Furthermore, we show that ZK MIA remains highly effective even when the adversary only knows a small fraction (10%) of their target's location history. This demonstrates that effective MIAs can be performed by realistic adversaries, highlighting the need for strong DP protection.

KEYWORDS

Location data, Membership Inference Attack, Synthetic Data

1 INTRODUCTION

Human mobility and location data are widely used across many important domains, such as epidemiology [23, 25], humanitarian response [75], and finance [27], as they offer insights into movement and density patterns. However, many people are concerned about the extensive collection of personal location data [29, 73, 79], particularly since this data may provide information regarding a person's social, economic, and political life [22].

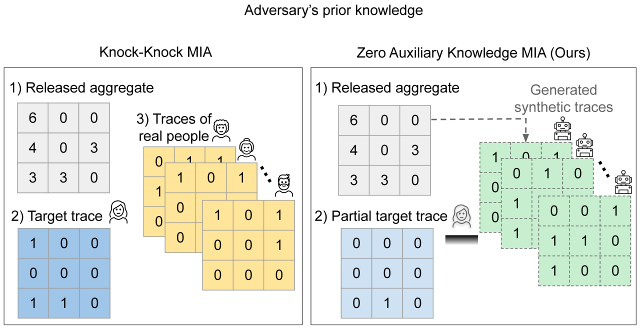


Figure 1: Adversary's prior knowledge in the previous work, Knock-Knock MIA [57] (left), and our work, Zero Auxiliary Knowledge MIA (right). The ZK adversary does not require knowledge of location traces of real people to run the MIA.

Individual-level location datasets have been shown to be highly vulnerable to re-identification attacks, due to the unicity and temporal consistency of people's mobility patterns [14, 70, 77]. To address these privacy concerns, data practitioners commonly use aggregate statistics, instead of individual-level records [3, 55, 74]. For example, the Public Health Agency of Canada studied citizens' movement during the COVID-19 pandemic, using aggregate location data from millions of mobile devices, provided by TELUS [52, 53]. British researchers conducted similar COVID-19 mobility analysis [35, 71], using aggregate location data obtained from O2 and Facebook. Because aggregate location data is often considered to be sufficiently de-identified [52], it is commonly sold by data brokers to interested parties [8, 62]. Notably, the U.S. government has been criticized for using commercial aggregate location data for law enforcement purposes and military intelligence [62]. Aggregate location data is also used in other sectors, such as urban design, to optimize public transit networks [36, 46, 50], and finance, to understand consumer behaviour [56, 59].

Motivation. As outlined in the E.U. Article 29 Working Party's guidance on anonymization techniques, aggregation reduces the risk of re-identification but does not eliminate all privacy risks [1]. In particular, aggregates may still be vulnerable to membership

[*] These authors contributed equally to this work.

Proceedings on Privacy Enhancing Technologies 2024(4), 80–101

https://doi.org/10.56553/popets-2024-0108

inference attacks (MIAs), whose goal is to infer if an individual's data was included in the data release, e.g. aggregate data. MIAs have become the de facto standard in privacy auditing due to their practical threat model and theoretical properties. From a practical perspective, a successful MIA is a direct privacy violation whenever participation in the data release is sensitive [42]. Furthermore, MIAs can be used as building blocks for other attacks, by first inferring a user's participation and then inferring their sensitive attributes. From a theoretical perspective, the success rate of an MIA is upper bounded following the application of differential privacy (DP) [18, 31, 76]. Hence, MIAs can be used as an auditing tool for DP implementations [32, 49]. Today, MIAs are widely used to assess the privacy risk of a broad range of data releases, including aggregate genetic data [28, 61], aggregate survey data [6], aggregate location data [51, 57], machine learning models [34, 49, 65] and synthetic data releases [24, 30, 44, 68].

MIAs pose an especially strong privacy threat on aggregate location data, since location data is often processed alongside sensitive attributes, such as socioeconomic status [71] and vaccination status [29]. In a notable example, a high-ranking priest resigned after being outed as homosexual by a radical group that matched his smartphone data with location data from Grindr, a popular dating app among the LGBTQ+ community [8]. It is therefore important to understand the practical risk that MIAs pose on aggregate location data, particularly by a realistic adversary, who only possesses information about their target.

The first and most prominent MIA on aggregate location data was proposed by Pyrgelis et al. [57]. Their "Knock-Knock" (KK) MIA works by training a binary classifier on a set of aggregates, wherein the adversary includes the target trace half of the time, and labels the aggregates accordingly. However, in addition to knowing the target trace, KK MIA requires the adversary to have access to a large auxiliary dataset of **individual-level traces** over the **same locations** and from a **similar population** as in the aggregate release. This is, when it comes to location data, a very strong assumption. This reliance on a strong adversary has led companies and practitioners to dismiss the risk posed by MIAs on location data. To the best of our knowledge, all previous works studying MIAs on aggregate location data require a similar auxiliary dataset [51, 58, 78].

Contributions. To assess the realistic privacy risk of releasing aggregate location data, we introduce the Zero Auxiliary Knowledge (ZK) MIA. ZK MIA is the first MIA on aggregate location data that does not require the adversary to have access to an auxiliary dataset. To remove this strong assumption, we develop a novel synthetic data-based approach, in which the adversary generates a reference dataset of synthetic traces, using only statistical parameters estimated from the aggregate. Training aggregates are then created using the synthetic reference. To account for privacy mechanisms applied to the release, we develop techniques to correct the parameter estimation for bias and noise, which enables ZK MIA to effectively attack privacy-aware aggregates as well. We also demonstrate that a paired sampling technique further improves MIA performance by isolating the contribution of the target trace within the high-dimensional aggregate. In the setting of ε-DP aggregate location data, we show that paired sampling enables MIAs to approach the worst-case ε-DP bound, offering a significant increase in performance to previous implementations.

We evaluate our Zero Auxiliary Knowledge (ZK) MIA against the state-of-the-art Knock-Knock (KK) MIA from Pyrgelis et al. [57] using two location datasets: i) a large-scale call detail record (CDR) dataset, and ii) the Milan Twitter dataset [67] from the Telecom Italia Big Data Challenge [5]. We apply the MIAs on raw and privacy-aware aggregates computed over 1000 users. Our results show that our ZK MIA closely matches the performance of KK MIA, without depending on extensive prior knowledge. On raw aggregates, both MIAs achieve 0.99 AUC on both datasets, suggesting that aggregation in itself is an ineffective safeguard. Both MIAs also surpass 0.9 AUC on both datasets under common privacy settings, including $\epsilon = 1$ event-level DP noise addition.

We further relax assumptions and show that the adversary does not need the full target trace for ZK MIA to succeed. Indeed, ZK MIA still achieved 0.84 AUC on the CDR dataset, with $\epsilon = 1$ event-level DP in place, when the adversary only knew a random 10% of the target trace.

After extensive evaluations across different privacy mechanisms, namely the suppression of small counts [9] and ε-DP noise addition [18], we argue that the commonly used ϵ-DP implementations on aggregate location data [16] do not protect against realistic privacy threats, such as our ZK MIA. We conclude that the only effective mitigation is the application of strong user level DP or user-day level DP guarantees, which is not yet a common practice [43, 50, 56, 59, 69].

2 DEFINITIONS AND THREAT MODEL

We formally define location traces and aggregates in Section 2.1 and overview aggregate-level privacy measures in Section 2.2. In Section 2.3 and 2.4, we outline the membership inference problem on aggregate location data and introduce the concept of a membership classifier. We present the threat model for our Zero Auxiliary Knowledge MIA and compare it against previous threat models for MIAs on location aggregates in Section 2.5. Table 1 of the Appendix contains a glossary of common terms.

2.1 Location Traces and Aggregates

Let $\mathcal{S} = \{s_1, ..., s_{|\mathcal{S}|}\}$ represent the set of all regions of interest (ROIs) where location data is collected. Similarly, $\mathcal{T} = \{t_1, ..., t_{|\mathcal{T}|}\}$ denotes the set of time intervals, also known as epochs, during which data collection occurs. In this paper, we assume that the geographic positions (i.e. approximate longitude and latitude) of the ROIs are known. For example, $\mathcal{S}$ may represent a set of square regions that partition a city into a grid, and $\mathcal{T}$ may represent contiguous hours over one month.

We focus on the scenario where location data of a set of users Ω is collected over the ROIs $\mathcal{S}$ and the epochs $\mathcal{T}$. We define the *location trace* L^u of a user $u \in \Omega$ as the set of geo-tagged and time-stamped visits (s, t) that u made within $\mathcal{S}$ during $\mathcal{T}$. We formally represent a user's location trace as the binary matrix

$$L^u_{s,t} = \begin{cases} 1 & \text{if user } u \text{ visited ROI } s \text{ during epoch } t \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

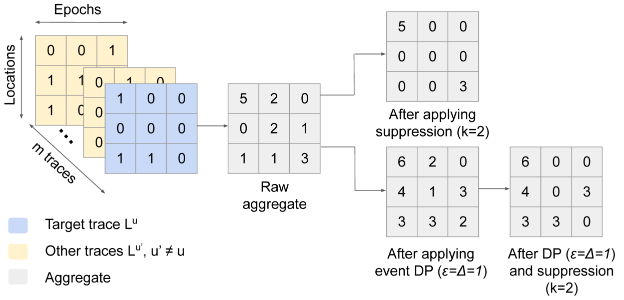


Figure 2: Example of how suppression of small counts and differential privacy may be applied to an aggregate with 3 ROIs (rows) and 3 epochs (columns).

Let $\mathcal{U} \subset \Omega$ be a group of m users whose location data is aggregated. We define an *aggregate* $A^{\mathcal{U}}$ to be the aggregate count statistics for $\mathcal{U}$ over $\mathcal{S} \times \mathcal{T}$. Formally, this is defined by the sum

$$A^{\mathcal{U}} = \sum_{u \in \mathcal{U}} L^u. \tag{2}$$

The entry $A^{\mathcal{U}}_{s,t}$ therefore corresponds to the number of users in $\mathcal{U}$ who visited ROI s during epoch t.

2.2 Privacy Measures on Location Aggregates

The data collector may be wary of the privacy risks of releasing the raw aggregate $A^{\mathcal{U}}$, and therefore apply privacy measures before releasing it.

2.2.1 *Differential Privacy.*

Differential privacy (DP) [18] is considered the gold standard for releasing information while protecting the privacy of individuals with formal guarantees. In essence, DP requires that the output of a computation over a dataset should not depend too much on the inclusion of any one record.

Definition 1 (ε-DP [18]). *A randomized algorithm M satisfies ε-DP if for all neighbouring datasets $D_1 \sim D_2$ (i.e., differing in exactly one record), and all possible outputs $S \subset Range(M)$:*

$$\Pr(M(D_1) \in S) \leq e^{\varepsilon} \Pr(M(D_2) \in S) \tag{3}$$

Thus, ε-DP limits the amount of information that can be inferred about individual records in the dataset, according to the privacy budget ε [18]. However, the privacy protection depends on what one considers as a "record", or *privacy unit*, when defining the neighbouring datasets. The most common definitions, in increasing level of privacy protection, are event-level, user-day level, and user-level DP (see Desfontaines [16] for an overview). The privacy unit for event-level DP is an individual data entry by any given user. For aggregate location data, this would be a single visit by a user to a ROI s during an epoch t. The privacy unit for user-day level would be all visits registered by any given user over a day. Finally, for user-level, the unit would be all visits in any given user's trace.

Randomised ε-DP mechanisms can be designed by adding noise sampled from the Laplace distribution [18]. $A^{\mathcal{U}}_{DP}(\varepsilon) = A^{\mathcal{U}} + Lap(\frac{\Delta}{\varepsilon})$ would satisfy (3) and be an ε-DP aggregation mechanism, where Δ is the global sensitivity, determined by the privacy unit.

In this paper, we assume the common practice of post-processing to ensure legitimate aggregate counts [21, 57, 58, 80]. Negative counts are set to 0, counts exceeding the group size m are set to m, and counts are rounded down to the nearest integer. These transformations will preserve ε-DP due to the post-processing theorem [17]. We note that the adversary can always apply these transformations themselves if the data collector does not do so already.

2.2.2 *Suppression of Small Counts (SSC).*

SSC is a privacy mechanism that aims to protect user privacy by hiding rare values. It has been frequently used across different types of datasets [11, 20, 58], including mobility datasets [3, 38]. Instead of releasing the raw aggregate $A^{\mathcal{U}}$, the data collector may choose a threshold $k \in \mathbb{N} \cup \{0\}$, and release the suppressed aggregate

$$A^{\mathcal{U}}_{SSC}(k)_{s,t} = \begin{cases} A^{\mathcal{U}}_{s,t} & \text{if } A^{\mathcal{U}}_{s,t} > k \\ 0 & \text{if } A^{\mathcal{U}}_{s,t} \leq k. \end{cases} \tag{4}$$

$A^{\mathcal{U}}_{SSC}(k)$ therefore contains the true count of users who visited a ROI s during epoch t as long as the count exceeds k. Lesser visited pairs (s, t) that record k or less visits are reported as 0 instead. Suppression can also be applied following ε-DP noise addition, such that (4) is applied on a noisy aggregate $A^{\mathcal{U}}_{DP}(\varepsilon)$. This produces $A^{\mathcal{U}}_{DP,SSC}(\varepsilon, k)$, an ε-DP aggregate whose final counts have been suppressed with threshold k. This transformation would preserve ε-DP due to post-processing [17], and may add a layer of complexity that mitigates attacks in practice.

2.3 Problem Formulation

We assume that the data collector releases aggregate count statistics $\overline{A}^{\mathcal{U}}$ over the ROIs $\mathcal{S}$ and the epochs $\mathcal{T}$, for the m users in the group $\mathcal{U}$. There are various cases depending on the privacy measures applied prior to release:

$$\overline{A}^{\mathcal{U}} = \begin{cases} A^{\mathcal{U}}, & \text{if the raw aggregate counts are released,} \\ A^{\mathcal{U}}_{DP}(\varepsilon), & \text{if only } \varepsilon\text{-DP is applied,} \\ A^{\mathcal{U}}_{SSC}(k), & \text{if only threshold } k \text{ SSC is applied,} \\ A^{\mathcal{U}}_{DP,SSC}(\varepsilon, k), & \text{if threshold } k \text{ SSC is applied after } \varepsilon\text{-DP.} \end{cases}$$

The goal of an adversary Adv performing an MIA on $\overline{A}^{\mathcal{U}}$ is to determine whether their target u^* contributed to $\overline{A}^{\mathcal{U}}$, inferring IN for $u^* \in \mathcal{U}$ and OUT for $u^* \notin \mathcal{U}$.

2.4 Membership Classifier

Given an aggregate release $\overline{A}^{\mathcal{U}}$ over m users, an adversary infers membership of the target u^* within the aggregation group $\mathcal{U}$ by using a binary membership classifier. Classifiers are commonly instantiated as machine learning models [57, 58, 78], but statistical models, like the log-likelihood function, have been applied as well [6, 28]. To train the classifier, Adv typically creates a balanced set of labeled size m training aggregates [51, 57, 58, 78]. Half of the aggregates include the target trace L^{u^*} and are labeled IN, and the other half are labeled OUT. Training the classifier will create a decision boundary in the underlying space of aggregate releases [7]. In the case of aggregate location data over ROIs $\mathcal{S}$ and epochs $\mathcal{T}$, the decision boundary is characterized by a hypersurface that partitions the matrix space $\mathbb{R}^{|\mathcal{S}|\times|\mathcal{T}|}$ into two sets.

2.5 Threat Model

In this section, we present our Zero Auxiliary Knowledge (ZK) MIA threat model. It is commonly assumed in MIAs across various domains that the adversary has access to an auxiliary dataset and complete knowledge of the target record [48, 57, 60, 65, 72, 76]. Our ZK threat model relaxes both assumptions by eliminating the need for an auxiliary dataset and allowing for only partial knowledge of the target trace.

For context, we also describe threat models of previous MIAs on aggregate location data. All threat models consider an adversary Adv, whose goal is to determine whether a specific target user u^* is included in the released aggregate $\overline{A}^{\mathcal{U}}$. The aggregate is computed across m users over ROIs $\mathcal{S}$ and epochs $\mathcal{T}$. We assume that the locations of the ROIs $\mathcal{S}$ are known.

Knock-Knock [57]: The adversary has an auxiliary dataset $Ref = \{L(u_1), ..., L(u_{|Ref|})\}$ of user traces, over the same locations and a similar population as the released aggregate. Ref has at least m traces, including the full target trace L^{u^*}.

LocMIA [78]: The adversary knows u^*'s social network and has an auxiliary dataset $Ref = \{L(u_1), ..., L(u_{|Ref|})\}$ of user traces, over the same locations and a similar population as the released aggregate $\overline{A}^{\mathcal{U}}$. Ref has at least m traces, including the traces of u^*'s friends, but not L^{u^*}.

Zero Auxiliary Knowledge (ours): The adversary knows a subset of the target u^*'s visits. Equivalently, the adversary knows a partial target trace $\tilde{L}^{u^*}$, such that $supp(\tilde{L}^{u^*}) \subset supp(L^{u^*})$.

KK MIA and LocMIA are reliant on the adversary's access to an extensive auxiliary dataset Ref. In particular, Adv samples individual traces from Ref to create training aggregates. These traces are also assumed to range over the same locations, and belong to a similar population as the traces aggregated in the release $\overline{A}^{\mathcal{U}}$, in order to properly train the membership classifier (Section 2.4). However, individual traces are known to be sensitive [14], and are unlikely to be made available, particularly when the data is aggregated as a privacy measure. Furthermore, because Ref must contain at least m traces, this assumption is impractical for even moderately sized aggregates. Although LocMIA removes prior knowledge about the target trace L^{u^*}, it must assume knowledge of u^*'s friends' traces to create a suitable proxy. More importantly, LocMIA still requires a large auxiliary dataset of individual traces.

In contrast, our Zero Auxiliary Knowledge adversary only requires that the adversary has knowledge about the target's location history. We emphasize that the adversary does not need to know the full trace L^{u^*}. Our threat model encompasses the case where only a few of the target's visits are known to the adversary. For example, the adversary may infer some of u^*'s visits from social media activity or direct observation.

3 RELATED WORK

MIAs on Aggregate Location Data. MIAs on aggregate location data have been shown to be successful on multiple location datasets [57, 58, 78], using a binary classifier to perform the inference task. The performance of the MIAs on small aggregates (< 500) is especially well-studied, as the influence of the target is easier to distinguish [57]. For example, the KK MIA by Pyrgelis et al. [57], achieved $AUC > 0.83$ when attacking size 100 aggregates across two different mobility datasets. LocMIA [78] is another MIA on aggregate location data, which removes prior knowledge about the target trace. Instead, LocMIA assumes access to social network information, and the traces of the target's friends, in order to construct a proxy for the target's real trace. However, both KK MIA and LocMIA crucially require the adversary to have access to a large auxiliary dataset to train the binary classifier. In contrast, our ZK MIA does not require any auxiliary dataset, and only requires partial information about the target trace (e.g. A random 10% proportion of their visits). ZK MIA therefore addresses the research gap of the MIA risk posed by a less knowledgeable attacker. The distinctions in prior knowledge are discussed in depth in Section 2.5. Our ZK MIA also features a novel approach, being the first MIA on aggregate location data to use synthetic trace generation.

Generation of Synthetic Location Traces. There are many techniques for generating synthetic location traces that capture realistic human mobility patterns [33, 37, 39–41, 54]. However, since our ZK MIA requires generating suitable synthetic traces without using additional information, this heavily limits the scope of applicable techniques. RNNs, GANs, and copulas have been used to generate synthetic traces that collectively approximate a real mobility dataset [39, 40, 54]. However, these techniques require real traces to train the model, which the ZK adversary does not have. Many of the state-of-the-art mobility models are also unsuitable because they simulate small-scale continuous trajectories (e.g. walks on campus) [37, 41]. In contrast, location aggregates typically comprise discrete traces over a metropolitan region. We therefore identified a probabilistic unicity model by Farzanehfar et al. [19], which requires only four statistical parameters to guide the synthetic generation. We demonstrate that we can non-trivially adapt this model for the ZK MIA. In particular, we develop methods to precisely estimate these parameters from the aggregates in order to produce realistic synthetic location aggregates.

MIAs with Reduced Auxiliary Data. Previous attempts have been made [10, 24, 60, 65, 72, 76] to relax the standard assumption of an adversary's access to an auxiliary dataset that has high statistical similarity with the attacked dataset, e.g. sampled from the same distribution [48, 65]. In the setting of machine learning (ML) models, where an MIA infers whether a record was a part of the ML model's training set, Shokri et al. [65] proposed an MIA without auxiliary data. Instead, they use synthetic data, which they generate using the ML model's confidence scores. Similarly, Salem et al. [60] trained an MIA using unrelated data, e.g., training on text data to attack an image model. These approaches require access to the ML model, and they train on features that are specific to ML models, such as the top-k confidence scores, which may be shared across ML models pertaining to different types of data. In contrast, the features used to train an MIA on aggregate location data explicitly depend on the

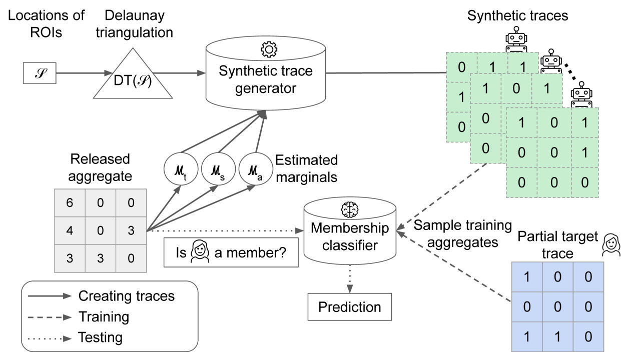


Figure 3: ZK MIA architecture: ***Adv*** **first creates synthetic traces, then uses them with the partial target trace to train the membership classifier, before predicting membership.**

specific regions, times, and population over which the aggregates are computed. In the context of synthetic generators, Guépin et al. [24] performed an MIA against synthetic data using the released synthetic dataset as the auxiliary dataset. However, in the case of aggregate location data, the release cannot be directly used as a reference dataset, since it does not contain individual records.

4 METHODOLOGY

We dedicate Sections 4.1-4.3 to explaining the synthetic-based methodology of our Zero Auxiliary Knowledge MIA. Section 4.4 explains the paired sampling mechanism on training aggregates, which boosts the performance of ZK MIA and KK MIA, as shown in Section 6.4.

4.1 Zero Auxiliary Knowledge MIA Framework

We implement the Zero Auxiliary Knowledge MIA as a binary classifier. However, whereas *Adv* uses their auxiliary dataset as a reference for creating training aggregates for KK MIA and LocMIA, *Adv* instead uses the reference of synthetic traces that they created from the released aggregate. Furthermore, if the full target trace L^{u^*} is not known, *Adv* may instead use a partial trace when creating the *IN* training aggregates. Figure 3 illustrates ZK MIA's overall attack architecture.

4.2 Generating Synthetic Traces from Aggregate Location Data

In order to generate synthetic traces for our Zero Auxiliary Knowledge MIA, we adapt a probabilistic mobility model [19]. Farzanehfar et al. [19] developed this model to reproduce unicity patterns in large populations. The model requires four statistical parameters, described below and illustrated in Figure 16 of the Appendix. Recall that for a discrete random variable X taking values in $x_1, ..., x_N$, its probability mass function (p.m.f.) $\mathcal{P} : \{x_1, ..., x_N\} \to [0, 1]$ maps each possible value to its corresponding probability.

(1) The marginal space distribution $\mathcal{P}_S : \mathcal{S} \to [0, 1]$ is a p.m.f. such that $\mathcal{P}_S(s)$ is proportional to the number of visits to ROI s by users in Ω across all epochs in $\mathcal{T}$.
(2) The marginal time distribution $\mathcal{P}_T : \mathcal{T} \to [0, 1]$ is a p.m.f. such that $\mathcal{P}_T(t)$ is proportional to the number of visits during epoch t by users in Ω across all ROIs in $\mathcal{S}$.
(3) The marginal activity distribution $\mathcal{P}_A$ models the total number of visits recorded within $\mathcal{S}$ during $\mathcal{T}$ by a user drawn from Ω.
(4) The Delaunay triangulation, denoted $DT(\mathcal{S})$, is a triangulation with vertices corresponding to the set of positions (longitude and latitude) of ROIs in $\mathcal{S}$. $DT(\mathcal{S})$ has the property that no vertex lies inside the circumcircle of any triangle in $DT(\mathcal{S})$ [15].

We note that the Delaunay triangulation $DT(\mathcal{S})$ is determined by the locations of the ROIs. Since the locations of the ROIs are assumed to be known (Section 2.5), $DT(\mathcal{S})$ can be immediately obtained from the release. We explain how the other statistical inputs, the three marginal distributions, can be approximated from the released aggregate in Section 4.3.

We now describe our procedure, adapted from Farzanehfar et al. [19], for generating synthetic traces using the four inputs. For each synthetic trace L^{syn}, we first sample the number of visits n_{visits} from the activity marginal $\mathcal{P}_A$. This determines the number of nonzero entries in the $\mathcal{S} \times \mathcal{T}$ matrix $L^{syn}_{s,t}$. Second, we sample an origin ROI s_0 from the space marginal $\mathcal{P}_S$, and a connected subgraph $C(s_0)$ from the Delaunay triangulation $DT(\mathcal{S})$, such that $s_0 \in C(s_0)$. $C(s_0)$ will correspond to the set of ROIs that may be visited in L^{syn}. This is done to emulate the natural tendency to move to and from the same proximate locations (e.g. home and work). Finally, we sample n_{visits} spatiotemporal visits (s, t) for which we set $L^{syn}_{s,t} = 1$. For each visit, s is sampled from the space marginal $\mathcal{P}_S$ restricted to $C(s_0)$, and t is sampled from the time marginal $\mathcal{P}_T$. All the sampling steps are independent.

We make two modifications of the original algorithm [19]. First, to avoid over-saturating unpopular regions, we sample the origin ROI s_0 according to the space marginal $\mathcal{P}_S$ rather than uniformly. Second, to allow users to visit multiple ROIs within the same epoch, we sample the epochs with replacement rather than without replacement. Our procedure is summarized in Algorithm 1 of the appendix.

4.3 Obtaining Accurate Marginals

In order to generate suitable synthetic traces for ZK MIA, *Adv* must estimate the marginal distributions $\mathcal{P}_S, \mathcal{P}_T, \mathcal{P}_A$, computed over the full population Ω, using the aggregate release $\overline{A}^{\mathcal{U}}$. In this section, we motivate and justify the techniques that we developed to obtain strong estimates $\widehat{\mathcal{P}}_S, \widehat{\mathcal{P}}_T, \widehat{\mathcal{P}}_A$. This task is especially challenging when privacy measures distort the aggregate data. We develop separate techniques to correct for bias in the case of SSC, and to correct for noise in the case of DP. The effects are shown in Figures 4 and 5 respectively. Algorithm 2 in the Appendix summarizes how we approximate all three marginals from $\overline{A}^{\mathcal{U}}$.

4.3.1 ***Estimating Space and Time Marginals.*** Suppose that the data collector releases the aggregate $\overline{A}^{\mathcal{U}}$, which may or may not have privacy measures. *Adv* can directly compute the empirical space and time marginals, which we denote $\widehat{\mathcal{P}}^0_S$ and $\widehat{\mathcal{P}}^0_T$, from the

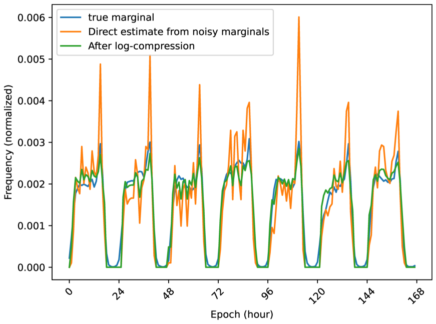


Figure 4: Log compression for SSC aggregates: SSC biases the estimate obtained from the aggregate by creating more extreme values. The true time marginal from the CDR dataset (plotted for the first week) is better approximated after the empirical estimate from the aggregate ($m = 1000$, $k = 1$) undergoes log compression $\log(1 + \gamma x)$ with γ chosen as in (8).

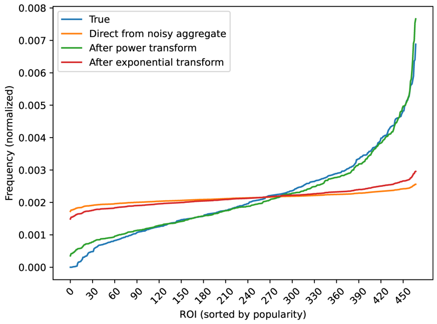


Figure 5: Power transformation for DP aggregates: DP noise compresses the estimate obtained from the aggregate. The true space marginal from the CDR dataset (organized by popularity) is better approximated after the empirical estimate from the aggregate ($m = 1000$, $\frac{\Delta}{\varepsilon} = 1$) undergoes power transformation x^p with p selected according to Algorithm 4.

released aggregate matrix $\overline{A}^{\mathcal{U}}$.

$$\widehat{\mathcal{P}}^0_S(s) = \frac{1}{\|\overline{A}^{\mathcal{U}}\|_1} \sum_{t=1}^{|\mathcal{T}|} \overline{A}^{\mathcal{U}}_{s,t} \tag{5}$$

$$\widehat{\mathcal{P}}^0_T(t) = \frac{1}{\|\overline{A}^{\mathcal{U}}\|_1} \sum_{s=1}^{|\mathcal{S}|} \overline{A}^{\mathcal{U}}_{s,t} \tag{6}$$

Raw Aggregate: In the case where the released aggregate provide the raw counts, i.e. $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}$, the empirical marginals tend to be highly accurate. An example is shown in Figure 13 in the Appendix. The accuracy of these estimates is intuitive because we expect the mobility patterns of an aggregation group $\mathcal{U}$ to resemble those of the population Ω. Thus, we set $\widehat{\mathcal{P}}_{\_} = \widehat{\mathcal{P}}^0_{\_}$ if the released aggregate is unmodified. We use the subscript $\_$ to indicate generality for both the space S and time T marginals.

Suppressed Aggregate: However, if the data collector applies SSC with threshold k, i.e. $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}_{SSC}(k)$, then this will systematically bias the empirical marginal $\widehat{\mathcal{P}}^0_{\_}$, because popular ROIs and epochs are more likely to evade suppression. It is therefore easy to see that suppression will reduce the observed probabilities of less popular entries and boost the probabilities of more popular entries.

To correct the bias, we flatten the empirical estimate $\widehat{\mathcal{P}}^0_{\_}$ by boosting low frequency counts and reducing high frequency counts. Upon the insight that $\widehat{\mathcal{P}}^0_{\_}$ can be likened to an audio signal, we adapt the logarithmic compression technique used to reduce dynamic range [45, 47]

$$x \longrightarrow \log\left(1 + \gamma x\right), \; x \geq 0, \tag{7}$$

where the scaling factor $\gamma \geq 0$ regulates the compression level [47]. In music signal processing, $x \geq 0$ corresponds to the intensity of a given frequency. In our case, $x \geq 0$ corresponds to probabilities within the empirical marginal $\widehat{\mathcal{P}}^0_{\_}$. We choose

$$\gamma(\widehat{\mathcal{P}}^0_{\_}) = \max_{x \in \widehat{\mathcal{P}}^0_{\_} : x > 0} \frac{1}{x} \tag{8}$$

to automatically parameterize γ based on the smallest observed non-zero probability. We therefore estimate $\widehat{\mathcal{P}}_{\_} = \log(1 + \gamma(\mathcal{P}^0_{\_}))$, where we omit the normalization constant.

We do not argue that our choice of method and parameter γ is optimal. However, the debiasing substantially improves the estimate, as shown in Figure 4, and it is done without additional information.

DP Aggregate. If ε-DP noise is added to each entry in the aggregate release, i.e. $\overline{A}^{\mathcal{U}}_{s,t} = A^{\mathcal{U}}_{DP}(\varepsilon)_{s,t} = A^{\mathcal{U}}_{s,t} + Lap(\frac{\Delta}{\varepsilon})$, then the noise will overpower the signal in the computation of the empirical marginals (Eq. 6). This follows from the fact that location aggregates are high-dimensional sparse matrices [58]. Therefore, conversely to the SSC case, the probabilities within the empirical marginals are compressed, since each probability is characterized mostly by thousands of independent noise samples. This effect is visualized for several different noise scales in Figure 17a of the Appendix. We also prove that under strong sparsity assumptions, $\widehat{\mathcal{P}}^0_S$ converges to the discrete uniform distribution on $\mathcal{S}$ as the number of epochs in the observation period $|\mathcal{T}| \rightarrow \infty$, in Theorem A.2 of the Appendix.

To correct the low variance of the observed probabilities, we propose the power transformation x^p with $p > 1$, followed by renormalization. It is easy to see that this will increase the variance since the probabilities are in $[0, 1]$. Automatically calibrating the power $p > 1$ is a delicate matter. To do so, we start with $p = 1$ and augment p gradually until the transformed distribution achieves the target variance σ^2. Without any prior knowledge, we consider the case where each probability is randomly drawn. Equivalently, each probablity x_i is sampled from $Unif(0, 1)$, and then renormalized so that the total probability is 1. Let p_i denote the probabilities after normalization, and $\bar{p}$ be the mean of the normalized probabilities. For the space marginal, the variance is

$$\sigma^2 = \frac{\sum_{i=1}^{|S|}(p_i - \bar{p})^2}{|S|}. \tag{9}$$

For the variance from the time marginal, we replace $|S|$ with $|T|$ in the above equation. The algorithm for selecting p is given in Algorithm 4 of the Appendix. Figure 5 shows that our automatically

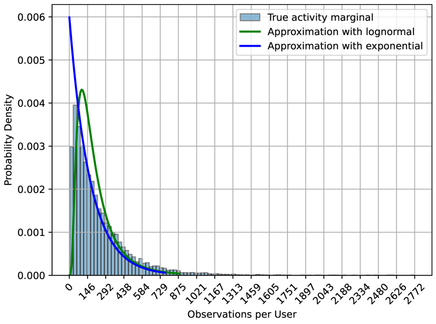


Figure 6: Exponential estimate for activity marginal: Once the mean number of visits is approximated, the activity marginal from the CDR dataset is best approximated by a lognormal distribution with an optimal skew parameter. However, it can also be approximated without additional parameters by an exponential distribution.

parameterized power transformation significantly improves the estimate. In contrast, the exponential transformation $(e^x - 1)/\gamma$, which is the inverse of the log compression function $\log(1 + \gamma x)$, fails to denoise because the inverse is inapplicable after considering normalization.

4.3.2 *Estimating the activity marginal.*

The released aggregate $\overline{A}^{\mathcal{U}}$ does not leak granular information about the activity marginal $\mathcal{P}_A$. However, *Adv* may obtain the empirical mean number of visits per user according to the released aggregate,

$$\widehat{\mu}_0 = \frac{1}{m} \sum_{s,t \in \mathcal{S} \times \mathcal{T}} \overline{A}^{\mathcal{U}}_{s,t}. \tag{10}$$

If the aggregate is raw, then we expect $\widehat{\mu}_0$ to be a strong estimate due to well-known regularity results about population-wide mobility activity [19, 63, 64]. In this case, *Adv* sets $\hat{\mu} = \widehat{\mu}_0$.

However, if either SSC or ε-DP is applied, then the estimate $\widehat{\mu}_0$ would fail. Algorithm 3 in the Appendix describes how *Adv* can obtain a better estimate $\hat{\mu}$. Given an aggregate release $\overline{A}^{\mathcal{U}}$ of size m, *Adv* can use $\widehat{\mathcal{P}}_S$ and $\widehat{\mathcal{P}}_T$ to iteratively improve their estimate, starting with μ_0. Given guess $\widehat{\mu}_n$, *Adv* creates a synthetic aggregate by generating m synthetic traces, parameterized by $\widehat{\mathcal{P}}_S, \widehat{\mathcal{P}}_T$ and $\widehat{\mathcal{P}_A} \sim \widehat{\mu}_n$. *Adv* may then apply the same privacy measures that were applied on $\overline{A}^{\mathcal{U}}$. $\widehat{\mu}_{n+1}$ is obtained by increasing or decreasing $\widehat{\mu}_n$ relative to the difference in counts with $\overline{A}^{\mathcal{U}}$.

Once $\hat{\mu}$ is obtained, *Adv* can simply pick $\widehat{\mathcal{P}_A} \sim \widehat{\mu}$, such that each synthetic trace has $\hat{\mu}$ visits. However, it is well known that human mobility activity follows a heavy-tailed distribution, i.e., a heavier tail than the exponential distribution. The best approximations are lognormal, beta, or power-law distributions [19, 63, 64].

It would be reasonable for *Adv* to use a heavy-tailed distribution with mean $\hat{\mu}$, but these distributions require a second parameter, e.g. skewness, to determine the distribution shape. *Adv* can use well-known parameters from other cities' datasets to complete the estimate [63, 64], but to ensure that *Adv* does not use additional knowledge, we assume that they use the sub-optimal estimate $\hat{\mathcal{P}}_A \sim Exp(\hat{\mu})$, as shown in Figure 6.

4.4 Paired Sampling for Training

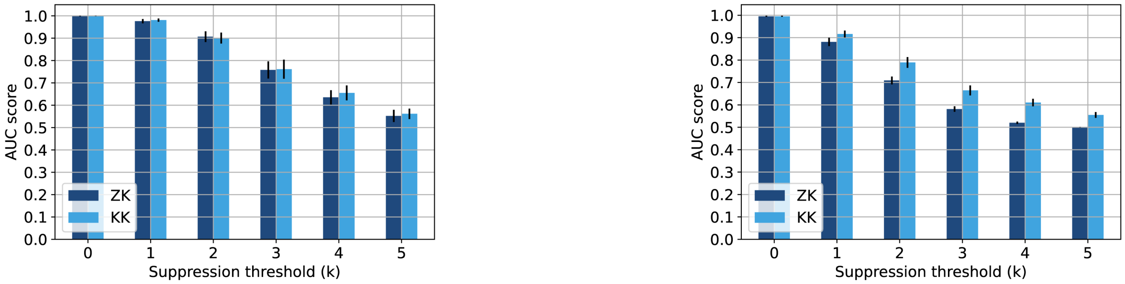
When MIAs target high-dimensional aggregate data, such as location data, the membership classifier must handle noise arising from thousands of entries, which are unrelated to the target record. For example, many of the *IN* training aggregates may coincidentally have high counts in entries that are absent in the target record. This would skew the decision boundary of the membership classifier, which may lead to false positives when testing. We may similarly obtain false negatives due to spurious patterns within the *OUT* training aggregates. These challenges are compounded by the implementation of privacy measures. For example, $\varepsilon-$DP would add noise to each entry. Given the dimensionality and nature of the aggregate data, spurious patterns will likely skew the decision boundary, even when hundreds or thousands of training aggregates are sampled. Given a fixed number of training samples, we demonstrate that the way in which the training set is sampled strongly influences the performance of the MIA. In particular, the sampling technique can guide the convergence of the decision boundary in order to prevent misclassification due to noise.

To the best of our knowledge, all previous MIAs on aggregate location data sampled their training set via independent random sampling [51, 57, 58, 78]. Training aggregates are created by independently sampling groups of m users from the population Ω and labeling them according to the target u^*'s presence.

On the one hand, independent sampling discourages overfitting to the training data by exposing the classifier to a wide variation of samples. On the other hand, independent sampling does nothing to prevent spurious patterns from distorting the decision boundary.

We propose a paired sampling technique to guide the convergence of the decision boundary. The idea is to use sampling to help the classifier identify the differential impact of the target record at the aggregate level. Paired sampling independently samples groups of $m - 1$ users from $\Omega \setminus \{u^*\}$. Then, an *IN* sample is created by adding u^* as the group's last member, and an *OUT* sample is created by adding another randomly selected user. The training set is therefore characterized by a set of *IN*/*OUT* pairs, which differ in exactly one record (the target's). If noise is added to aggregates prior to release, then *Adv* must inject the same noise sample ε to each paired sample, $A^{\mathcal{U}^{IN}}$ and $A^{\mathcal{U}^{OUT}}$, to ensure that the target's differential impact is preserved between each *IN*/*OUT* pair.

Paired sampling therefore actively encourages the membership decision boundary to be formed based on relevant criteria related to the target. It also discourages spurious decision boundaries because of the high degree of similarity between *IN*/*OUT* pairs. An extreme value in an aggregate entry from an *IN* sample will be matched with a similar value from its paired *OUT* sample, with high probability. However, we note that using paired sampling effectively halves the training variation compared to independent sampling. Our experiments in Section 6.4 demonstrate that paired sampling outperforms independent sampling across all tested settings of ε-DP noise addition. Hence, guiding the decision boundary towards relevant membership criteria often takes precedence over maximizing training variation. For completeness, we note that while we



(a) CDR dataset

(b) Milan

Figure 7: Mean AUC scores with standard error for ZK and KK on size 1000 aggregates with various suppression thresholds

developed, studied, and named paired sampling independently, we later found that a similar idea was used in Bauer and Bindschaedler [6] but had not been compared to independent sampling nor used elsewhere in the literature so far to the best of our knowledge.

5 EXPERIMENTAL SETUP

To evaluate the efficacy of our ZK MIA, we compare it against the state-of-the-art Knock Knock (KK) MIA [57, 58] using aggregated location data from two different datasets.

5.1 Datasets.

In this section, we describe the two location datasets used for evaluating the MIAs, and discuss ethical considerations of the data collection and usage.

CDR: The first dataset, which we refer to as "CDR", is a private dataset, shared with us by Flowminder [2] for the purpose of this research. The raw dataset comprises timestamped and geo-tagged call records of approximately 11,000 mobile phone users within a Latin American metropolitan area. The observation period is June 2021, with epochs defined by the 720 hourly timeslots. The ROIs are defined by the service regions of approximately 500 cellular antenna towers within the metropolitan area, which spans $\sim 150\text{km}^2$. The users were selected such that they registered at least one visit per week, to omit users who changed SIM cards, and such that the majority of their visits are within the region, to ensure that they are residents. 50 target users for the MIAs were randomly selected by Flowminder. A histogram of the number of visits over the target traces is plotted in Figure 14a in the Appendix.

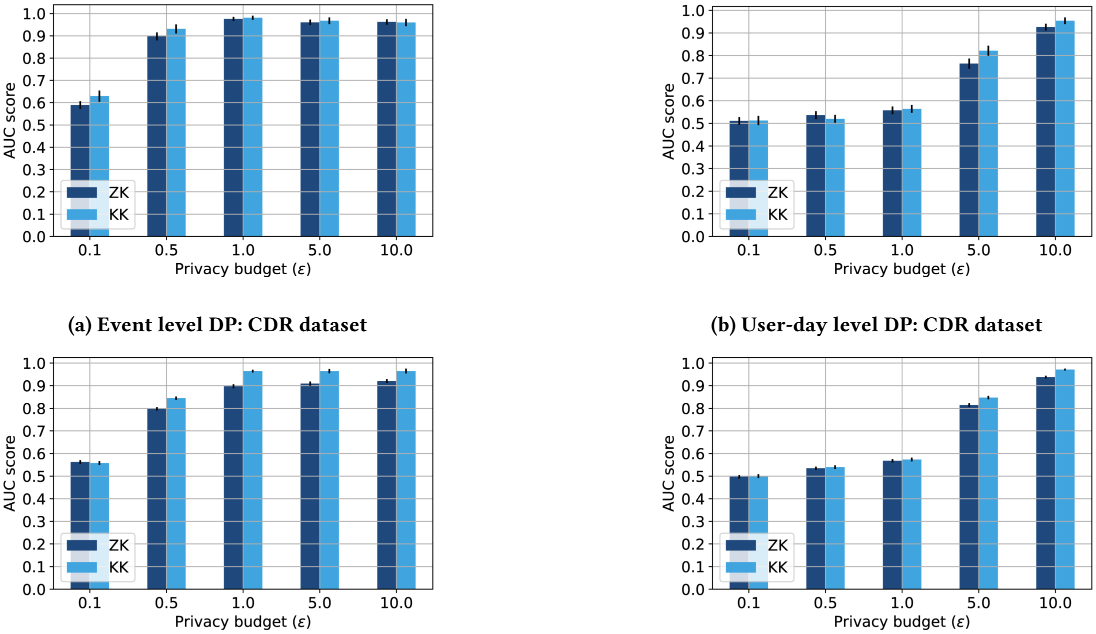
Milan: The second dataset is the Milan Social Pulse dataset [67], made publicly available as part of the Telecom Italia Big Data Challenge [5]. This dataset comprises timestamped and geo-tagged tweets from 4840 mobile phone users within the Milano region. The ROIs are defined by a grid of 100 points, each with an approximate area of 256 km^2. We consider the location data from the first week of data, yielding 168 hourly epochs. We do not delete any users from the dataset prior to aggregation. We randomly select 50 targets among users who tweeted at least 10 times during the observation period. A histogram of the number of visits over the target traces is plotted in Figure 14b in the Appendix.

Ethical Considerations: Because of the sensitivity of location data, we did not access raw individual-level data, and instead collaborated with Flowminder (FM) to develop a privacy-preserving data-sharing pipeline for the purpose of this research [13]. More specifically, data sharing was restricted to pre-computed aggregate matrices (labeled according to target membership, computed by FM on the data provider server) and 50 target traces randomly chosen by FM. To further mitigate the privacy risk, the ~ 500 ROI and 720 epoch indices were randomly permuted in the shared aggregate and target trace matrices, according to a mapping known only by FM. This random permutation relabeled the space and time indices, enabling us to test the MIAs without knowing the true times or locations. The graphs of the marginal statistics (see Figures 4, 5, 6) were plotted by FM and shared with us. All the data shared by FM with us is subject to a research contract between FM and our institution and was kept on our segregated server. The Milan dataset, derived from geo-tagged tweets, remains publicly available, and was only used for the purpose of testing the MIAs.

5.2 MIA Implementation

We perform a fair comparison between KK MIA and ZK MIA by training the binary membership classifier using the same parameters and architecture. This also helps us isolate the effect of removing auxiliary data on performance. We use a Logistic Regression binary classifier with default hyperparameters and $\mathbb{L}_1$ regularization, implemented with *sklearn*. The number of training groups $n_{train} = 400$ matches previous implementations [57, 58], and the groups are selected using paired sampling, unless specified otherwise. We additionally fine tune the decision boundary using $n_{val} = 100$ balanced independently sampled validation groups. We flatten the aggregates and feed them directly into the classifier as a vector, without any processing, such as PCA or feature extraction. Finally, as done in Pyrgelis et al. [57], *Adv* applies the same privacy measures to training and validation aggregates, if the released aggregate is privacy-aware.

Knock-Knock. To implement the Knock-Knock MIA, we provide *Adv* with a reference group Ref of 5000 real user traces (including the target trace L^{u^*}) when attacking the larger CDR dataset. We set $|Ref| = 2500$ for the Milan dataset. We note that this significantly surpasses previous reference sizes ($|Ref| = 1100$) implemented by Pyrgelis et al. [57].



(a) Event level DP: CDR dataset

(b) User-day level DP: CDR dataset

(c) Event level DP: Milan

(d) User-day level DP: Milan

Figure 8: Mean AUC scores with standard error for ZK and KK on size 1000 aggregates with various privacy units and budgets ε.

Zero Auxiliary Knowledge. Our Zero Auxiliary Knowledge MIA is structurally identical to KK (PS), but ZK has a synthetic reference Ref, rather than a set of real traces. This reference is created according to the methodology detailed in Section 4. Setting $|Ref| = 5000$ allows for a direct comparison of the functionality of synthetic traces to real ones for the purpose of the MIA. However, we remark that capping the number of synthetic traces at 5000 is an artificial restriction, since *Adv* may generate arbitrarily more synthetic traces and achieve better performance, as shown in Figure 12 of the Appendix. By default, we assume full access to the target trace L^{u^*}. This assumption is relaxed for experiment 6.3.

5.3 Evaluation

Default Experimental Parameters. We randomly select $n_{targets} = 50$ targets from the dataset for evaluation. These targets are re-used for each experiment. Furthermore, in each experiment, $n_{test} = 100$ *independently sampled* balanced test aggregates are created for each target, and shared across both MIAs to ensure that the test sets are identical. As done in [57], the test aggregate user groups are sampled from a disjoint set of users to the Knock-Knock adversarial reference Ref, plus the target u^*. This corresponds to roughly $11,000 - 5000 = 6000$ user traces for CDR and $5000 - 2500 = 2500$ user traces for Milan.

We perform all experiments on size $m = 1000$ aggregates, which matches the largest aggregate size tested in Pyrgelis et al. [57] and exceeds the largest aggregate size tested in Zhang et al. [78] (800). We do not vary m since the relationship between aggregate group size and MIA effectiveness has already been documented extensively [57, 58, 78]. We perform the Knock-Knock and Zero Auxiliary Knowledge MIAs in this setting under different privacy measures. We also perform experiments such that *Adv* only knows a fraction p_{u^*} of the target trace L^{u^*}, but by default, we assume that they know the full trace, i.e. $p_{u^*} = 1$.

Evaluation Metrics. In the past, MIAs on aggregate location data have been primarily evaluated using the area under the ROC curve (AUC score) as a metric [57, 78]. For this reason, and its suitability for assessing the strength of a binary classifier, we use the mean AUC over all targets as our primary metric. However, we also include the mean attack accuracy over all targets as a secondary metric, listing these scores in Section C of the Appendix.

6 EXPERIMENTAL RESULTS

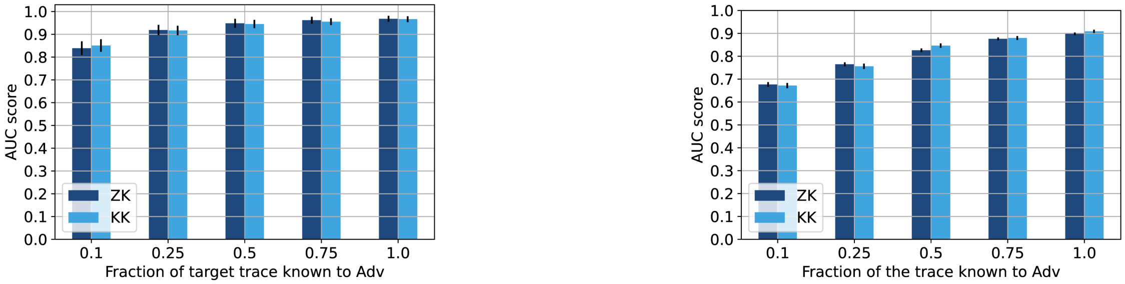
6.1 Against Suppression of Small Counts

We first compare the performances of KK and ZK on aggregates whose counts have been suppressed according to threshold k. We apply SSC with thresholds $k \in \{0, 1, 2, 3, 4, 5\}$ on test aggregates of size $m = 1000$. We note that the case $k = 0$ corresponds to releasing a raw aggregate. We remark that there is a trivial rule that sufficiently determines non-membership in this special case.

Rule ($k = 0$): If u^* visits ROI s during epoch t and no users in the aggregation group $\mathcal{U}$ visit (s, t), then u^* cannot be in $\mathcal{U}$, i.e.,

$$\exists s \in \mathcal{S}, t \in \mathcal{T} : A^{\mathcal{U}}_{s,t} = 0 \wedge L^{u^*}_{s,t} = 1 \implies u^* \notin \mathcal{U}$$

We therefore incorporate this rule when $k = 0$, such that both MIAs first check if the released aggregate elicits the contradiction. If so, we immediately predict *OUT*. Otherwise, we train, validate, and test the classifier as usual. The rule is invalid for $k > 0$, since it would predict *OUT* whenever u^* has a visit to a suppressed entry.



(a) CDR dataset

(b) Milan

Figure 9: Mean AUC scores with standard error for ZK and KK on size 1000 aggregates with $\varepsilon = 1$ event DP protection and $k = 1$ suppression for varying fractions of the known target trace.

Results. Figure 7 shows that membership inference is a trivial task when applied to raw aggregates ($k = 0$). Both ZK and KK achieve near perfect AUC (≥ 0.99) on both datasets. This implies that aggregation is not an effective privacy mechanism in itself to protect high-dimensional location data from MIAs by weak or strong adversaries.

The results for $k > 0$ reveal two general patterns. First, ZK compares closely with KK across different levels of SSC. On the CDR dataset, ZK's AUC stays within 0.02 of KK for each k. We observe slightly worse results on the Milan dataset, but ZK still stays within 0.9 AUC of KK for each k. Second, there is a monotonic decrease in performance when the threshold k is increased. For $k = 5$, the AUC is always less than 0.55. This follows from the fact that suppression reduces the amount of available information, and 99% of all nonzero entries are suppressed by $k = 5$, as shown in Figure 15 of the Appendix. Therefore, although SSC eventually mitigates the MIAs, it may come at the cost of destroying virtually all utility.

Both MIAs perform worse on the Milan dataset compared to the CDR dataset. This is expected because we observe 6 times less data per target in the Milan dataset, as shown in Figure 14. Indeed, people generally tweet less often than they text and call. ZK MIA is more affected by dataset sparsity, given its dependence on marginal distribution estimates, which become less reliable in sparser datasets.

6.2 Against ε-DP Noise Addition

Informed by practical applications of DP [16], we consider event level and user-day level to be the privacy units, and we vary the privacy budget $\varepsilon \in \{0.1, 0.5, 1.0, 5.0, 10.0\}$ for each unit.

Event Level DP. An event is equivalent to a visit by a user to $(s, t) \in \mathcal{S} \times \mathcal{T}$. To offer privacy protection over an event, we set the global sensitivity $\Delta = 1$. ϵ-DP is then ensured by adding $Lap(\frac{1}{\epsilon})$ noise to each count in the aggregate matrix.

User-day Level DP. In order to protect each user's daily contributions without adding excessive noise, it is common to restrict user contributions prior to aggregation to achieve a smaller global sensitivity Δ [3, 26]. We analysed daily activity distributions, and had them preprocessed such that a user may only contribute up to $\Delta = 20$ visits in any given day for CDR, and $\Delta = 10$ visits in any given day for Milan. ϵ-DP at the user-day level is then ensured by adding $Lap(\frac{\Delta}{\epsilon})$ noise to each count in the aggregate matrix.

Results. Figure 8 shows that ZK MIA matches KK MIA across all tested DP settings. Indeed, ZK maintained a mean AUC within 0.06 of KK (PS) across each of the 10 privacy settings for both datasets. KK and ZK notably succeeded for many of the tested privacy budgets ε, particularly in the event level setting. Indeed, we observed $AUC \geq 0.9$ for both MIAs whenever the noise scale $\frac{\Delta}{\epsilon} \leq 2$ for the CDR dataset, and $\frac{\Delta}{\epsilon} \leq 1$ for the Milan dataset. These settings are in line with many real-life applications [16, 38]. Conversely, user-day level DP with privacy budget $\varepsilon \leq 1.0$ effectively reduced both MIAs to an AUC below 0.55. We discuss the significance of these results with respect to practical mitigations in Section 7.

6.3 Partial Knowledge of the Target Trace

We now relax the assumption that Adv knows the full target trace L^{u^*}. This is in line with our ZK threat model, and we expand KK MIA for this setting to be able to compare methods. To simulate a weaker adversary, we suppose that Adv only knows a subset of the target u^*'s visits. We assume that Adv only knows a random fraction $p_{u^*} \in \{0.1, 0.25, 0.5, 0.75, 1.0\}$ of the trace L^{u^*}. The number of retained visits is rounded up to the next integer to prevent cases where Adv knows 0 visits. For example, if a target has 4 visits and $p_{u^*} = 0.1$, then this would correspond to Adv knowing 1 visit. This partial trace is used instead of the full trace when creating IN training and validation aggregates. The full trace L^{u^*} is still used for IN test aggregates.

We perform this experiment in the setting where the data collector applies event-level DP with $\epsilon = 1$, followed by $k = 1$ suppression. We choose this setting for a couple of reasons. First, to study the degradation of the MIAs with decreasing information about the target, we choose a setting where Adv would succeed given the full target trace. Previous experiments revealed that $\epsilon = 1$-DP at the event level and $k = 1$ suppression were not effective in preventing the MIAs by themselves, as the MIAs achieved AUC > 0.97 on the CDR dataset and AUC > 0.9 for Milan. Second, we combine the two defense mechanisms to see if suppression has an observable mitigation effect when applied following DP noise addition. By the post-processing property of DP, this would not alter the theoretical performance bound. However, zeroing all counts ≤ 1 might add a layer of complexity that affect MIAs in practice.

Results. First, we note that applying $k = 1$ SSC on top of $\varepsilon = 1.0$ event level DP has an insignificant effect on the MIAs. For the full target trace, we continue to observe AUC > 0.97 on the CDR dataset and AUC > 0.9 on the Milan dataset. The only MIA with a noticeable decline was KK MIA on the Milan dataset, which dropped from 0.97 AUC to 0.9 AUC.

Although decreasing the fraction of the target trace known to the adversary from 1 to 0.1 decreases the performance of the MIAs, the corresponding degradation is relatively gradual. All AUCs are captured within a range of 0.13 on the CDR dataset, and within a range of 0.22 on the Milan dataset. Even the lowest observed AUC by ZK MIA on the CDR dataset (0.84 when 10% of the target trace is known) achieves high discrimination. We note that the 50 targets in both datasets have a wide variation in trace size, as shown in Figure 14 of the Appendix. For some targets, *Adv* will only know one of the target's visits, whereas for others, they will still know dozens and be able to infer membership easily. Interestingly, we note that knowing a single visit from a target trace can still train a classifier that is better than random. For one CDR target with 9 visits in their full trace, we observed ZK achieve an AUC of 0.660 across 100 aggregates when only 1 random visit was known. Although far from perfect, we found this surprising, as it shows that even a single visit by the target can inform an MIA against a noisy aggregate over 1000 users.

6.4 Paired Sampling vs. Independent Sampling

In this section, we study the performance of KK MIA and ZK MIA when we vary the sampling technique used for creating their training aggregates, i.e., paired sampling (PS) or independent sampling (IS). To test this, we consider the implementation of user-day $\varepsilon - DP$ on the Milan dataset across the privacy budgets $\varepsilon = 1, 2, 3, ..., 10$ for all four possible MIAs: KK (PS), KK (IS), ZK (PS), ZK (IS).

Results. From Figure 10, we see that the paired sampling MIA always outperforms its independent sampling equivalent across all privacy budgets $\varepsilon = 1, 2, 3, ..., 10$. Paired sampling provides the largest boost when the inference task is challenging but not intractable. In particular, we notice a few striking examples of ZK (PS) drastically outperforming ZK (IS) in the middle of the graph. For $\epsilon = 4$, ZK (IS) is basically random ($AUC = 0.54$), yet simply switching to paired sampling enables the classifier to achieve an AUC of 0.83.

The improvement achieved by switching from independent sampling to paired sampling is less significant for KK in this experiment. This suggests that using training aggregates sampled from a reference dataset of real traces may introduce less randomness to the membership classifier's decision boundary, compared to when we use a synthetic reference. This is intuitive because of ZK's probabilistic generation method, which relies on sampling from three different estimated distributions. However, the MIAs have indistinguishable performance when both attacks use paired sampling, with the difference in AUC always staying within 0.02 of one another across the 10 privacy settings. This suggests that paired sampling effectively eliminates the noise contributed by coincidental patterns in random entries, and enables the membership classifier to form a suitable decision boundary.

7 DISCUSSION

We first provide a critical analysis of the experimental results, followed by a discussion of mitigation strategies and their practicality. We then consider limitations in our methods and evaluations, before discussing how our methodology may be generalized to MIAs beyond the setting of aggregate location data.

7.1 Analysis of Results: Practical Risk of MIA

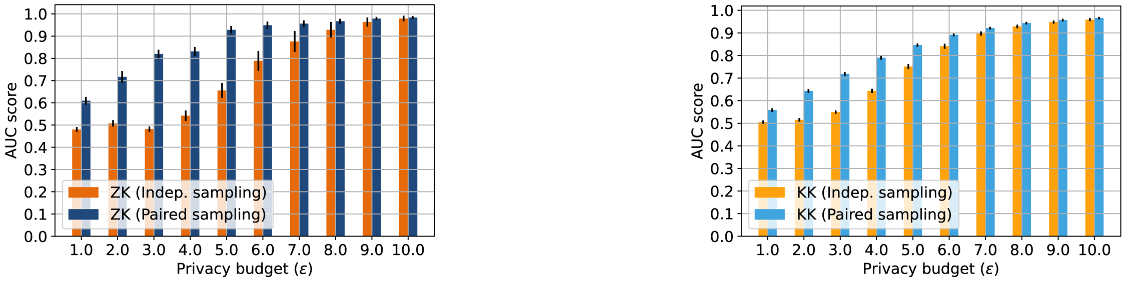
In Section 6, we observed that our ZK MIA achieves approximately the same performance as the KK MIA across all experiments, and that both MIAs performed effectively across a range of common privacy settings. This has several important implications.

First, the ZK MIA significantly increases the attack surface of aggregate location data, since no auxiliary dataset is needed for the MIA. Although previous MIAs on aggregate location data have been successful, the strong assumption of a large auxiliary dataset prevents these attackers from attempting the MIA in most real-life cases. The auxiliary dataset comprises sensitive user-level information collected from the same dataset that is being aggregated. However, aggregation is applied to prevent the release of personal information. Moreover, *Adv* would be restricted by the size of their auxiliary dataset, since they would not be able to perform MIAs on aggregates computed over more users than there are in their reference. In contrast, we demonstrated that our Zero Auxiliary Knowledge *Adv* can create an arbitrary number of synthetic traces upon seeing the released aggregate, without an auxiliary dataset. This offers *Adv* the flexibility to attack aggregates of any size. Section D.1 in the Appendix shows results for KK and ZK MIA across aggregates of size $m = 100, 500, 1000, 2000, 3000$. Figure 12 of the Appendix also shows that the ZK *Adv* can boost their own performance up to diminishing marginal returns, simply by generating more traces.

We have also shown that MIAs on aggregate location data are more powerful than previously known. By incorporating paired sampling for training, we have demonstrated more effective MIA results on size 1000 aggregates than previously reported [58], particularly when protected by differential privacy. Our results therefore demonstrate that MIAs on aggregate location data are easily performed without auxiliary data, more effective than previously believed, and that common privacy measures fail to protect against the risk.

7.2 Proposed Mitigations

Our results show that aggregated location data requires more stringent privacy safeguards to protect against MIAs. This is an inherently challenging task because our ZK MIA was able to succeed by using the aggregate to estimate where the population moves (space marginal), when the population moves (time marginal), and how frequently the population moves (activity marginal). However, location aggregates naturally leak this information. In fact, much of its utility is derived from these marginal statistics. Therefore, while one can mitigate the ZK MIA by perturbing the aggregate to



(a) Zero Auxiliary Knowledge

(b) Knock-Knock

Figure 10: Mean AUC scores with standard error for ZK and KK on size 1000 aggregates from the Milan dataset across different privacy budgets ε under user-day level ε-DP, for MIAs using independent sampling vs. paired sampling.

the point where the basic mobility patterns of its population are unrecoverable, doing so may also destroy the aggregates' utility.

In light of these results, we advise that data practitioners be mindful of the parameters that they select for $\epsilon-$DP, because DP does not guarantee sufficient protection from MIAs if the parameters are chosen too loosely. Since data practitioners often prioritize utility, it is common to pick more relaxed parameters for the privacy unit (e.g. event or user-day instead of user) and budget ε (e.g. $\epsilon > 1$). For example, Kohli et al. [38] studied $\epsilon \in \{0.1, 0.5, 1.0\}$ at the event level in the context of aggregate O-D mobility matrices, Facebook used $\epsilon = 0.45$ at the event level when collecting data about URLs shared on the site, and Apple uses ϵ between 2 and 16 at the user-day level when collecting IOS data [16]. Recall that ZK and KK achieved $AUC \geq 0.9$ on the CDR dataset whenever the noise scale $\frac{\Delta}{\epsilon}$ was 2 or less. This corresponds to $\epsilon \geq 0.5$ for event level DP and $\epsilon \geq 10$ for user-day level DP with up to $\Delta = 20$ daily visits. ZK and KK therefore both achieved high discrimination on privacy settings that are in line with many real-life applications.

However, we do observe DP mitigating the MIAs when we pick sufficiently strict parameters. For example, no MIA achieved better than random performance for $\epsilon = 0.5$ in the user-day setting. We also note that we did not evaluate using the user level setting, which would achieve the strongest privacy protection. We note that the suitability of privacy parameters depends on the desired utility and sensitivity of the dataset. A stricter parameter choice is particularly relevant if the aggregate is publicly released and/or pertains to sensitive data.

Although ε-DP always offers privacy guarantees, our experimental results emphasize the importance of picking appropriate parameters. In particular, we observed that event-level DP was largely ineffective in preventing MIAs from both strong and weak adversaries. We instead encourage the use of user-day or user-level DP with carefully selected privacy budgets ε to mitigate the practical threat of an MIA.

7.3 Limitations

We have so far taken the Knock-Knock MIA to refer to the Subset of Locations setting [57]. We now address why we do not consider the Knock-Knock Participation in Past Groups [57] threat model in this paper. Under the Participation in Past Groups setting, the adversary has access to a set of past aggregates $\{\overline{A}^{\tilde{\mathcal{U}}_1}, ..., \overline{A}^{\tilde{\mathcal{U}}_N}\}$, collected over the same ROIs $\mathcal{S}$ as the released aggregate $\overline{A^{\mathcal{U}}}$. Moreover, *Adv* is assumed to know the membership status of the target u^* in each of these aggregates. That is, *Adv* knows whether or not $u^* \in \mathcal{U}_i$ for all $i = 1, ..., N$. This last assumption is crucial because *Adv* directly uses $\{\overline{A}^{\tilde{\mathcal{U}}_1}, ..., \overline{A}^{\tilde{\mathcal{U}}_N}\}$ as their training data for the membership classifier in this setting. This is unrealistic for multiple reasons. First, to train an effective membership classifier, there would need to be hundreds of labeled aggregates to have sufficient training data. More importantly, there would be no reason for the membership status of an individual within an aggregate to be released in practice. We argue that the only plausible scenario in which *Adv* would know the membership status of each aggregate is if they created the aggregates themselves. This reduces to the Subset of Locations setting that we have assumed in this paper.

In terms of limitations for our ZK MIA, recall that the Delaunay triangulation of the ROIs, $DT(\mathcal{S})$, is the only non-probabilistic parameter used to generate synthetic traces for the ZK MIA. The triangulation only depends on the locations of the ROIs, which we have so far assumed to be shared as part of the aggregate release (Section 2.5). We believe this to be a realistic assumption, as omitting the locations of the ROIs would strongly diminish the utility of aggregate location data. Nonetheless, there might exist cases where released location aggregates do not relay the positions of the ROIs. For example, Google binned ROIs into categories, ex. restaurants, parks, and hospitals, when publicly releasing their mobility report during COVID https://www.google.com/covid19/mobility/. In this setting, the adversary would proceed without knowing where the ROIs are situated with respect to one another. The privacy risk under this setting is not known, and we identify it as an area of future research. Similarly, there might exist cases where the adversary knows the ROIs that were visited by the target (ex. home and work), but not the visitation times. We show in Appendix D.3 that only knowing the visited ROIs substantially reduces the effectiveness of both MIAs.

ZK MIA also requires that we estimate statistical parameters from the released aggregate. It may be difficult to estimate these precisely if the aggregate size is small or if the collected location data is not regular. However, we still observe strong performance by ZK MIA on both datasets for small aggregate sizes (see Appendix D.1).

Furthermore, aggregate location data collected over large metropolitan populations are known to obey high regularity across different cities and time periods. These patterns include log-normal activity distributions [19, 63, 64] and periodic "circadian rhythm" time marginals [12, 19, 64, 66]. This suggests that our statistical parameter estimation should be highly transferable across sufficiently regular datasets. However, we acknowledge that there are scenarios where the observed population is not regular (e.g. taxi drivers).

7.4 Generalizations to MIAs in other Settings

In this paper, we have proposed a new methodology to perform membership inference attacks on aggregate data, by training the attack on synthetic records, generated from the released aggregate. We believe that this approach can be adapted for MIAs in settings beyond aggregate location data. Our methodology can be broken down into two main steps: 1) extracting noise-less global statistics from the released aggregate, 2) use these statistics to create individual-level records to train the MIA.

In the setting of location data, the relevant statistics pertain to the mobility trends of large-scale human populations [12, 19, 63, 64, 64, 66], and individual location [37, 39–41, 54] which have both been well established in the literature. This facilitates both steps of our methodology, as we know in advance what location data should look like at both the global and individual level.

Although the trends will be distinct from aggregate location data, aggregate releases for other types of data will generally reveal global statistics. For instance, categorical tabular data is modeled by discrete random variables, whereas location data is modeled by continuous random variables, and approximated by high-dimensional discrete data. Our methods for denoising and debiasing statistics from differentially private and suppressed aggregates are however not specific to location data, and should generalize to other data releases. Regarding the second step, using the statistics to create individual records for training, the probabilistic method used for our ZK MIA, drawing from the Delaunay triangulation and the relevant marginal distributions, is partially specific to location data. One would thus need to carefully consider the statistical properties of the type of data to create high quality individual records.

8 CONCLUSION

Aggregate location data is widely shared and used by governments [29, 53, 62], companies [3, 4, 26], and researchers [35, 38, 71] because of its insights into human behaviour and its presumed security against reidentification.

In this paper, we demonstrated that aggregate location data is susceptible to MIAs by realistic adversaries, who only know some of their target's location history. With ZK MIA, we introduced the first MIA on aggregate location data that does not require an auxiliary dataset. We accomplished this by generating appropriate synthetic traces, using statistics that are estimated from the released aggregate. We also equipoed our parameter estimation with techniques that automatically correct for bias and noise from popular privacy mechanisms like suppression of small counts and ε-DP noise.

We then showed that MIAs on aggregate location data are significantly improved by incorporating a paired sampling technique, which helps isolate the effect of the target trace within a high dimensional aggregate. Hence, the vulnerability of aggregate location data is further heightened by these improved attacks.

Our evaluations over two large datasets demonstrate that, despite the absence of an auxiliary dataset, ZK MIA performs as well as the state-of-the-art KK MIA, with both MIAs achieving high discrimination when commonly used privacy settings are applied. ZK MIA remains effective in realistic privacy settings, even when only a small fraction (10%) of the target trace is known. These results emphasize the need for strict differential privacy guarantees on released aggregate location data.

Taken together, our findings show that membership inference attacks are not merely a theoretical privacy threat posed by unrealistically strong adversaries, but also a realistic threat to contend with in practice.

ACKNOWLEDGMENTS

Ana-Maria Cretu did most of her work while she was at Imperial College London and was partially funded by the Agence Française de Développement via the Flowminder Foundation.

The authors would like to sincerely thank the Flowminder team for their support with this work, in particular Galina Veres and James Harrison for their help in designing a secure data-sharing pipeline to test the MIAs on the CDR dataset. The authors would like to further acknowledge Cyril Miras for his early work on MIAs on aggregate location data, including the implementation of the baseline rule MIA. Finally, the authors would like to thank the anonymous reviewers and shepherd for their feedback on the paper.

REFERENCES

[1] (2014). Article 29 data protection working party. opinion 05/2014 on anonymisation techniques. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

[2] (2024). Flowminder website. https://www.flowminder.org/.

[3] Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., et al. (2020). Google covid-19 community mobility reports: anonymization process description (version 1.1). *arXiv preprint arXiv:2004.04145*.

[4] Apple, D. (2017). Learning with privacy at scale. *Apple Machine Learning Journal*, 1(8).

[5] Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., Antonelli, F., Vespignani, A., Pentland, A., and Lepri, B. (2015). A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data*, 2(1):1–15.

[6] Bauer, L. A. and Bindschaedler, V. (2020). Towards realistic membership inferences: The case of survey data. In *Annual Computer Security Applications Conference*, pages 116–128.

[7] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

[8] Boorstein, M. and Kelly, H. (2023). Colorado catholic group bought app data that tracked gay priests. *The Washington Post*.

[9] Chen, B.-C., Kifer, D., LeFevre, K., Machanavajjhala, A., et al. (2009). Privacy-preserving data publishing. *Foundations and Trends® in Databases*, 2(1–2):1–167.

[10] Crețu, A.-M., Guépin, F., and de Montjoye, Y.-A. (2021). Correlation inference attacks against machine learning models. *arXiv preprint arXiv:2112.08806*.

[11] Cretu, A.-M., Houssiau, F., Cully, A., and de Montjoye, Y.-A. (2022). Querysnout: Automating the discovery of attribute inference attacks against query-based systems. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 623–637.

[12] Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J.-C., Huens, E., Van Dooren, P., Smoreda, Z., and Blondel, V. D. (2013). Exploring the mobility of mobile phone users. *Physica A: statistical mechanics and its applications*, 392(6):1459–1473.

[13] de Montjoye, Y.-A., Gambs, S., Blondel, V., Canright, G., De Cordes, N., Deletaille, S., Engø-Monsen, K., Garcia-Herranz, M., Kendall, J., Kerry, C., et al. (2018). On the

privacy-conscientious use of mobile phone data. *Scientific data*, 5(1):1–6.

[14] de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1):1–5.

[15] Delaunay, B. et al. (1934). Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7(793-800):1–2.

[16] Desfontaines, D. (2021). A list of real-world uses of differential privacy.

[17] Dwork, C., Kohli, N., and Mulligan, D. (2019). Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2).

[18] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer.

[19] Farzanehfar, A., Houssiau, F., and de Montjoye, Y.-A. (2021). The risk of re-identification remains high even in country-scale location datasets. *Patterns*, 2(3):100204.

[20] Gadotti, A., Houssiau, F., Rocher, L., Livshits, B., and De Montjoye, Y.-A. (2019). When the signal is in the noise: Exploiting diffix's sticky noise. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1081–1098.

[21] Ge, Q. and Fukuda, D. (2016). Updating origin–destination matrices with aggregated data of gps traces. *Transportation Research Part C: Emerging Technologies*, 69:291–312.

[22] Georgiadou, Y., de By, R. A., and Kounadi, O. (2019). Location privacy in the wake of the gdpr. *ISPRS international journal of geo-information*, 8(3):157.

[23] Grantz, K. H., Meredith, H. R., Cummings, D. A., Metcalf, C. J. E., Grenfell, B. T., Giles, J. R., Mehta, S., Solomon, S., Labrique, A., Kishore, N., et al. (2020). The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology. *Nature communications*, 11(1):4961.

[24] Guépin, F., Meeus, M., Cretu, A.-M., and de Montjoye, Y.-A. (2023). Synthetic is all you need: removing the auxiliary data assumption for membership inference attacks against synthetic data. *arXiv preprint arXiv:2307.01701*.

[25] Hara, Y. and Yamaguchi, H. (2021). Japanese travel behavior trends and change under covid-19 state-of-emergency declaration: Nationwide observation by mobile phone location data. *Transportation Research Interdisciplinary Perspectives*, 9:100288.

[26] Herdağdelen, A. and Dow, A. (2021). Protecting privacy in facebook mobility data during the covid-19 response (2020). *URL https://research. fb. com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response*.

[27] Holmes, A., Byrne, A., and Rowley, J. (2013). Mobile shopping behaviour: insights into attitudes, shopping process involvement and location. *International Journal of Retail & Distribution Management*, 42(1):25–39.

[28] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., and Craig, D. W. (2008). Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4(8):e1000167.

[29] Hope, C. (2021). Millions 'unwittingly tracked' by phone after vaccination to see if movements changed.

[30] Houssiau, F., Jordon, J., Cohen, S. N., Daniel, O., Elliott, A., Geddes, J., Mole, C., Rangel-Smith, C., and Szpruch, L. (2022). Tapas: A toolbox for adversarial privacy auditing of synthetic data. *arXiv preprint arXiv:2211.06550*.

[31] Humphries, T., Oya, S., Tulloch, L., Rafuse, M., Goldberg, I., Hengartner, U., and Kerschbaum, F. (2023). Investigating membership inference attacks under data dependencies. In *2023 IEEE 36th Computer Security Foundations Symposium (CSF)*, pages 473–488. IEEE.

[32] Jagielski, M., Ullman, J., and Oprea, A. (2020). Auditing differentially private machine learning: How private is private sgd? *Advances in Neural Information Processing Systems*, 33:22205–22216.

[33] Jahromi, K. K., Zignani, M., Gaito, S., and Rossi, G. P. (2016). Simulating human mobility patterns in urban areas. *Simulation Modelling Practice and Theory*, 62:137–156.

[34] Jayaraman, B. and Evans, D. (2019). Evaluating differentially private machine learning in practice. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1895–1912.

[35] Jeffrey, B., Walters, C. E., Ainslie, K. E., Eales, O., Ciavarella, C., Bhatia, S., Hayes, S., Baguelin, M., Boonyasiri, A., Brazeau, N. F., et al. (2020). Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with covid-19 social distancing interventions was high and geographically consistent across the uk. *Wellcome Open Research*, 5.

[36] Kakakhel, S. (2022). Optimising urban planning with location intelligence. Quadrant Blog. Accessed: 2024-03-07.

[37] Karagiannis, T., Le Boudec, J.-Y., and Vojnović, M. (2007). Power law and exponential decay of inter contact times between mobile devices. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 183–194.

[38] Kohli, N., Aiken, E., and Blumenstock, J. (2023). Privacy guarantees for personal mobility data in humanitarian response. *arXiv preprint arXiv:2306.09471*.

[39] Kulkarni, V. and Garbinato, B. (2017). Generating synthetic mobility traffic using rnns. In *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery*, pages 1–4.

[40] Kulkarni, V., Tagasovska, N., Vatter, T., and Garbinato, B. (2018). Generative models for simulating mobility trajectories. *arXiv preprint arXiv:1811.12801*.

[41] Lee, K., Hong, S., Kim, S. J., Rhee, I., and Chong, S. (2009). Slaw: A new mobility model for human walks. In *IEEE INFOCOM 2009*, pages 855–863. IEEE.

[42] Li, N., Qardaji, W., Su, D., Wu, Y., and Yang, W. (2013). Membership privacy: A unifying framework for privacy definitions. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 889–900.

[43] Martínez-Durive, O. E., Mishra, S., Ziemlicki, C., Rubrichi, S., Smoreda, Z., and Fiore, M. (2023). The netmob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography. *arXiv preprint arXiv:2305.06933*.

[44] Meeus, M., Guepin, F., Creţu, A.-M., and de Montjoye, Y.-A. (2023). Achilles' heels: vulnerable record identification in synthetic data publishing. In *European Symposium on Research in Computer Security*, pages 380–399. Springer.

[45] Miguel Alonso, B. D. and Richard, G. (2004). Tempo and beat estimation of musical signals. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain*.

[46] Morgan, M. and Lovelace, R. (2021). Travel flow aggregation: Nationally scalable methods for interactive and online visualisation of transport behaviour at the road network level. *Environment and Planning B: Urban Analytics and City Science*, 48(6):1684–1696.

[47] Müller, M. (2015). Logarithmic compression. https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3S1_LogCompression.html.

[48] Nasr, M., Shokri, R., and Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE symposium on security and privacy (SP)*, pages 739–753. IEEE.

[49] Nasr, M., Songi, S., Thakurta, A., Papernot, N., and Carlin, N. (2021). Adversary instantiation: Lower bounds for differentially private machine learning. In *2021 IEEE Symposium on security and privacy (SP)*, pages 866–882. IEEE.

[50] O2 (2019). O2 transport smart steps product sheet. https://static-www.o2.co.uk/sites/default/files/2019-04/o2-transport-smart-steps-product-sheet.pdf. [Online].

[51] Oehmichen, A., Jain, S., Gadotti, A., and de Montjoye, Y.-A. (2019). Opal: High performance platform for large-scale privacy-preserving location data analytics. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 1332–1342. IEEE.

[52] Office of the Privacy Commissioner of Canada (2023). Investigation into the collection and use of de-identified mobility data in the course of the covid-19 pandemic. Accessed: 2023-09-14.

[53] Oli, S. (2021). Canada's public health agency admits it tracked 33 million mobile devices during lockdown. *National Post*, 24.

[54] Ouyang, K., Shokri, R., Rosenblum, D. S., and Yang, W. (2018). A non-parametric generative model for human trajectories. In *IJCAI*, volume 18, pages 3812–3817.

[55] Popa, R. A., Blumberg, A. J., Balakrishnan, H., and Li, F. H. (2011). Privacy and accountability for location-based aggregate statistics. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 653–666.

[56] Precisely (2024). Placeiq movement. https://www.precisely.com/product/precisely-placeiq/placeiq-movement. Accessed: 2024-02-13.

[57] Pyrgelis, A., Troncoso, C., and De Cristofaro, E. (2017). Knock knock, who's there? membership inference on aggregate location data. *arXiv preprint arXiv:1708.06145*.

[58] Pyrgelis, A., Troncoso, C., and De Cristofaro, E. (2020). Measuring membership privacy on aggregate location time-series. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(2):1–28.

[59] SafeGraph (2024). Enrich pois with aggregated transaction data. https://www.safegraph.com/products/spend. Accessed: 2024-02-13.

[60] Salem, A., Zhang, Y., Humbert, M., Berrang, P., Fritz, M., and Backes, M. (2018). Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models. *arXiv preprint arXiv:1806.01246*.

[61] Sankararaman, S., Obozinski, G., Jordan, M. I., and Halperin, E. (2009). Genomic privacy and limits of individual detection in a pool. *Nature genetics*, 41(9):965–967.

[62] Savage, C. (2021). Intelligence analysts use u.s. smartphone location data without warrants, memo says. *The New York Times*. Available at https://www.nytimes.com/2021/01/22/us/politics/dia-surveillance-data.html.

[63] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., and González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246.

[64] Seshadri, M., Machiraju, S., Sridharan, A., Bolot, J., Faloutsos, C., and Leskove, J. (2008). Mobile call graphs: beyond power-law and lognormal distributions. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 596–604.

[65] Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pages 3–18. IEEE.

[66] Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, 327(5968):1018–1021.

[67] SpazioDati and di Milano, D. P. (2015). Social Pulse - Milano.

[68] Stadler, T., Oprisanu, B., and Troncoso, C. (2022). Synthetic data–anonymisation groundhog day. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1451–1468.

[69] Telus (2024). Telus Insights Location API. https://docs.insights.telus.com/. [Online].

[70] Tournier, A. J. and de Montjoye, Y.-A. (2022). Expanding the attack surface: Robust profiling attacks threaten the privacy of sparse behavioral data. *Science Advances*, 8(33):eabl6464.

[71] Trasberg, T. and Cheshire, J. (2023). Spatial and social disparities in the decline of activities during the covid-19 lockdown in greater london. *Urban Studies*, 60(8):1427–1447.

[72] Truex, S., Liu, L., Gursoy, M. E., Yu, L., and Wei, W. (2019). Demystifying membership inference attacks in machine learning as a service. *IEEE Transactions on Services Computing*, 14(6):2073–2089.

[73] Van Zoonen, L. (2016). Privacy concerns in smart cities. *Government Information Quarterly*, 33(3):472–480.

[74] Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., and Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, 42:625–646.

[75] Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C., and Ukkusuri, S. V. (2022). Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, 94:101777.

[76] Yeom, S., Giacomelli, I., Fredrikson, M., and Jha, S. (2018). Privacy risk in machine learning: Analyzing the connection to overfitting. In *2018 IEEE 31st computer security foundations symposium (CSF)*, pages 268–282. IEEE.

[77] Zang, H. and Bolot, J. (2011). Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 145–156.

[78] Zhang, G., Zhang, A., and Zhao, P. (2020). Locmia: Membership inference attacks against aggregated location data. *IEEE Internet of Things Journal*, 7(12):11778–11788.

[79] Zhou, T. (2017). Understanding location-based services users' privacy concern: An elaboration likelihood model perspective. *Internet Research*, 27(3):506–519.

[80] Zhu, K., Fioretto, F., and Van Hentenryck, P. (2022). Post-processing of differentially private data: A fairness perspective. *arXiv preprint arXiv:2201.09425*.

APPENDIX

Table 1: Glossary of notations.

Notation	Definition
$\mathcal{S}$	Set of regions of interests (ROIs)
$\mathcal{T}$	Set of epochs in observation period
Ω	Set of all users in the dataset
L^u	Location trace of user $u \in \Omega$ over $\mathcal{S} \times \mathcal{T}$
$\mathcal{U}$	Aggregation group of users sampled from Ω
$A^{\mathcal{U}}$	Raw aggregate count matrix in $\mathcal{S} \times T$ over users in $\mathcal{U}$
$A^{\mathcal{U}}_{DP}(\varepsilon)$	An ε-DP aggregate
$A^{\mathcal{U}}_{SSC}(k)$	An aggregate with counts $\leq k$ suppressed
$A^{\mathcal{U}}_{DP,SSC}(\varepsilon, k)$	An ε-DP aggregate with counts $\leq k$ suppressed
$\overline{A}^{\mathcal{U}}$	The released aggregate count matrix
m	Number of users in the aggregation group $\mathcal{U}$
u^*	Target drawn from full population, $u^* \in \Omega$
Adv	Adversary performing MIA on u^*

A SUPPLEMENTARY PROOFS

Definition 2. *(Oracle average count) Given a raw aggregate $A^{\mathcal{U}}$, we define the oracle average count function $\mu_S : \mathcal{S} \to \mathbb{R}^+$ as*

$$\mu_S(s) = \lim_{|\mathcal{T}| \to \infty} \frac{\sum_{t=1}^{|\mathcal{T}|} A^{\mathcal{U}}_{s,t}}{|\mathcal{T}|}. \tag{11}$$

Letting $|\mathcal{T}| \to \infty$ corresponds to extending the observation period indefinitely. Thus, $\mu_S(s)$ represents the expected number of users $\mathcal{U}$ who visit ROI s at a randomly selected epoch, given infinite location data over the ROIs $\mathcal{S}$.

Table 2: Default experiment parameters.

Default value	Definition
$n_{train} = 400$	Number of training aggregates
$n_{val} = 100$	Number of validation aggregates
$n_{test} = 100$	Number of test aggregates
$n_{target} = 50$	Number of targets
$m = 1000$	Aggregate size
$\|Ref\| = 5000$ (CDR), 2500 (Milan)	Traces in Adv's real (KK) or synthetic (ZK) reference
$p_{u^*} = 1$	Fraction of L^{u^*} known by Adv

Definition 3. *(Strong sparsity) We say that $A^{\mathcal{U}}$ is strongly sparse if*

$$\mu_S(s) = 0 \ \forall s \in \mathcal{S} \tag{12}$$

Equivalently, $\sum_{t=1}^{|\mathcal{T}|} A^{\mathcal{U}}_{s,t} \in o(|\mathcal{T}|) \ \forall s \in \mathcal{S}$. This is a strong assumption, as it implies that the visitation rate to each ROI decreases at a sublinear rate.

Lemma A.1. *Given a fixed geographic region in which location data is collected,*

$$\lim_{|\mathcal{S}| \to \infty} \frac{\sum_{s=1}^{|\mathcal{S}|} A^{\mathcal{U}}_{s,t}}{|\mathcal{S}|} = 0 \tag{13}$$

Proof. $\sum_{s=1}^{|\mathcal{S}|} A^{\mathcal{U}}_{s,t}$ corresponds to the number of users who registered a visit during epoch t. Letting $|\mathcal{S}| \to \infty$ corresponds to increasing creating finer regional partitions within the fixed geographic region. $\sum_{s=1}^{|\mathcal{S}|} A^{\mathcal{U}}_{s,t}$ is invariant to increasing $|\mathcal{S}| \to \infty$, since the same users are observed over the same time. It follows that $\lim_{|\mathcal{S}| \to \infty} \frac{\sum_{s=1}^{|\mathcal{S}|} A^{\mathcal{U}}_{s,t}}{|\mathcal{S}|} = 0$. □

Theorem A.2. *(Convergence of empirical marginals to uniform distribution under ε-DP) Let $\Delta > 0$ be the global sensitivity and suppose that ε-DP is applied on an aggregate release $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}_{DP}(\varepsilon)$ with post-processed non-negative counts. If the original raw counts $A^{\mathcal{U}}$ are strongly sparse, then the empirical space and time marginals, $\mathcal{P}^0_S$ and $\mathcal{P}^0_T$, each converge to discrete uniform distributions:*

- $\widehat{\mathcal{P}}^0_S \to Unif(\mathcal{S})$ *in distribution as* $|\mathcal{T}| \to \infty$
- $\widehat{\mathcal{P}}^0_T \to Unif(\mathcal{T})$ *in distribution as* $|\mathcal{S}| \to \infty$

Proof. We first consider $\widehat{\mathcal{P}}^0_S$. It suffices to show that as $\epsilon \to 0$, $\widehat{\mathcal{P}}^0_S(s_0) \xrightarrow{\text{a.s.}} \frac{1}{|\mathcal{S}|}$ for each $s_0 \in \mathcal{S}$.

Let $\epsilon > 0$ and let $b = \frac{\Delta}{\epsilon}$. Recall that ε-DP with post-processed non-negative counts is obtained by $\overline{A}^{\mathcal{U}}_{s,t} = (A^{\mathcal{U}}_{s,t} + L_{s,t}(b)) \vee 0$, where $A^{\mathcal{U}}_{s,t}$ is the true number of visits by users in $\mathcal{U}$ to (s, t) and

$\{L_{s,t} \sim Lap(b) : s \in \mathcal{S}, t \in \mathcal{T}\}$ are i.i.d Laplacian noise samples (Section 2.2.1). By definition,

$$\begin{aligned}
\widehat{\mathcal{P}}^0_S(s_0) &= \frac{\sum_{t=1}^{|\mathcal{T}|} \overline{A}^{\mathcal{U}}_{s_0,t}}{\sum_{s=1}^{|\mathcal{S}|}\sum_{t=1}^{|\mathcal{T}|} \overline{A}^{\mathcal{U}}_{s,t}} \\
&= \frac{\sum_{t=1}^{|\mathcal{T}|}\left((A^{\mathcal{U}}_{s_0,t} + L_{s_0,t}(b)) \vee 0\right)}{\sum_{s=1}^{|\mathcal{S}|}\sum_{t=1}^{|\mathcal{T}|}\left((A^{\mathcal{U}}_{s,t} + L_{s,t}(b)) \vee 0\right)} \\
&= \frac{\frac{\sum_{t=1}^{|\mathcal{T}|}\left((A^{\mathcal{U}}_{s_0,t}+L_{s_0,t})\vee 0\right)}{|\mathcal{T}|}}{\sum_{s=1}^{|\mathcal{S}|}\frac{\sum_{t=1}^{|\mathcal{T}|}\left((A^{\mathcal{U}}_{s,t}+L_{s,t})\vee 0\right)}{|\mathcal{T}|}}.
\end{aligned}$$

We now express $(A^{\mathcal{U}}_{s,t} + L_{s,t}) \vee 0 = L_{s,t} \vee 0 + X_{s,t}$, for some $X_{s,t}$, in order to apply Lemma A.3 later. Since $A^{\mathcal{U}}_{s,t} \geq 0$, there are three cases:

$$X_{s,t} = \begin{cases} A^{\mathcal{U}}_{s,t}, & \text{if } L_{s,t} \geq 0 \\ A^{\mathcal{U}}_{s,t} + L_{s,t}, & \text{if } L_{s,t} < 0 \text{ and } (A^{\mathcal{U}}_{s,t} + L_{s,t}) \vee 0 > 0 \\ 0, & \text{if } L_{s,t} < 0 \text{ and } (A^{\mathcal{U}}_{s,t} + L_{s,t}) \vee 0 = 0 \end{cases}$$

We therefore have

$$\widehat{\mathcal{P}}^0_S(s) = \frac{\frac{\sum_{t=1}^{|\mathcal{T}|} X_{s_0,t}}{|\mathcal{T}|} + \frac{\sum_{t=1}^{|\mathcal{T}|} L_{s_0,t}\vee 0}{|\mathcal{T}|}}{\sum_{s=1}^{|\mathcal{S}|}\left(\frac{\sum_{t=1}^{|\mathcal{T}|} X_{s,t}}{|\mathcal{T}|} + \frac{\sum_{t=1}^{|\mathcal{T}|} L_{s,t}\vee 0}{|\mathcal{T}|}\right)}. \tag{14}$$

By sparsity, for each $s \in \mathcal{S}$

$$\frac{\sum_{t=1}^{|\mathcal{T}|} A^{\mathcal{U}}_{s,t}}{|\mathcal{T}|} \to 0 \text{ as } |\mathcal{T}| \to \infty \tag{15}$$

Also, by the Strong Law of Large Numbers, since $\{L_{s,t}\}$ are i.i.d., and $\mathbb{E}[Lap(b)] = 0$, we have

$$\frac{\sum_{t=1}^{|\mathcal{T}|} L_{s,t}}{|\mathcal{T}|} \xrightarrow{\text{a.s.}} 0 \text{ as } |\mathcal{T}| \to \infty$$

By linearity,

$$\frac{\sum_{t=1}^{|\mathcal{T}|} A^{\mathcal{U}}_{s,t} + L_{s,t}}{|\mathcal{T}|} \xrightarrow{\text{a.s.}} 0 \text{ as } |\mathcal{T}| \to \infty$$

Hence, in all three possible cases,

$$\frac{\sum_{t=1}^{|\mathcal{T}|} X_{s,t}}{|\mathcal{T}|} \xrightarrow{\text{a.s.}} 0 \text{ as } |\mathcal{T}| \to \infty$$

This allows us to simplify

$$\widehat{\mathcal{P}}^0_S(s) \xrightarrow{\text{a.s.}} \frac{\frac{\sum_{t=1}^{|\mathcal{T}|} L_{s_0,t}\vee 0}{|\mathcal{T}|}}{\sum_{s=1}^{|\mathcal{S}|}\frac{\sum_{t=1}^{|\mathcal{T}|} L_{s,t}\vee 0}{|\mathcal{T}|}}.$$

Since for all s, t, $L_{s,t} \vee 0 \sim Lap(b) \vee 0$, Lemma A.3 implies $\mathbb{E}[L_{s,t} \vee 0] = \frac{b}{2}$. Hence, by the Strong Law of Large Numbers,

$$\frac{\sum_{t=1}^{|\mathcal{T}|} L_{s,t} \vee 0}{|\mathcal{T}|} \xrightarrow{\text{a.s.}} \frac{b}{2} \text{ as } |\mathcal{T}| \to \infty$$

Finally, for any set of ROIs $\mathcal{S}$, and any $s \in \mathcal{S}$,

$$\widehat{\mathcal{P}}^0_S(s) \xrightarrow{\text{a.s.}} \frac{\frac{b}{2}}{\sum_{s=1}^{|\mathcal{S}|}\frac{b}{2}} = \frac{b}{|\mathcal{S}|b} = \frac{1}{|\mathcal{S}|},$$

A symmetric argument proves $\widehat{\mathcal{P}}^0_T \to Unif(\mathcal{T})$ in distribution as $|\mathcal{S}| \to \infty$, using Lemma A.1 instead of strong sparsity. □

Remark. We note that strong sparsity is assumed in Eq. (14) to prove that $\frac{\sum_{t=1}^{|\mathcal{T}|} X_{s,t}}{|\mathcal{T}|} \xrightarrow{\text{a.s.}} 0$ as $|\mathcal{T}| \to \infty$. Although we expect the oracle average count $\mu_S(s)$ to be very small for most $s \in \mathcal{S}$, due to the sparsity of aggregate location data, it is unlikely to observe $\mu_S = 0$ for real data. Substituting $\mu_S(s)$ in place of 0 in Eq. (14) will not yield the uniform probability $\widehat{\mathcal{P}}^0_S(s) \xrightarrow{\text{a.s.}} \frac{1}{|\mathcal{S}|}$, but it will be a close approximation, provided that $\frac{\Delta}{\epsilon} >> \mu_S$ and that the number of epochs is large.

In practice, fixed dimensions for S and T will prevent the empirical marginals from completely converging to the uniform distribution. This is demonstrated for different noise scales on the Milan dataset (which has $|S| = 100$ and $|T| = 168$) in Figure 17.

Lemma A.3. *Suppose that $Y \sim L \vee 0$, with $L \sim Lap(b)$. Then, Y has mean*

$$\mathbb{E}[Y] = \frac{b}{2}$$

Proof. Let $L \sim Lap(b)$. Then, its probability density function (pdf) f_L is given by

$$f_L(x) = \frac{1}{2b}e^{\frac{|x|}{b}} \text{ for } x \in \mathbb{R}$$

which is symmetric about $x = 0$. Hence, $P(X \leq 0) = \frac{1}{2}$. It follows that $Y = X \vee 0$ has the pdf f_Y

$$f_Y(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{\delta(x)}{2}, & \text{for } x = 0 \\ \frac{1}{2b}e^{-\frac{x}{b}}, & \text{for } x > 0 \end{cases}$$

where $\delta(x)$ is the Dirac delta function representing the accumulated probability mass at zero. We then evaluate

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{1}{2b}\int_0^\infty xe^{-\frac{x}{b}}dx \\
&= \frac{1}{2b}\left(-bxe^{-\frac{x}{b}}\Big|_0^\infty + b\int_0^\infty e^{-\frac{x}{b}}dx\right) \\
&= \frac{1}{2b}\left(-b^2e^{-\frac{x}{b}}\Big|_0^\infty\right) \\
&= \frac{1}{2b}\left(b^2\right) = \frac{b}{2}
\end{aligned}$$

□

B ALGORITHMS

In this section, we present the main algorithms required to generate synthetic traces from the released aggregate for our ZK MIA.

Algorithm 1 describes how we adapted the unicity model from Farzanehfar et al. [19] to generate synthetic traces for ZK MIA. We note that the procedure for generating a synthetic trace can also be interpreted as running a Markov chain $\{X_i : i = 1, ..., n_{visits}\}$ over the state space of spatiotemporal pairs $(s, t) \in C(s_0) \times \mathcal{T}$ with transition probabilities to $(s', t') \in C(s_0) \times \mathcal{T}$ proportional to the product of the pmfs $\mathcal{P}_S(s')\mathcal{P}_T(t')$.

Algorithm 2 estimates the three marginal probability distributions required to run Algorithm 1: the space marginal $\mathcal{P}_S$, the time marginal $\mathcal{P}_T$, and the activity marginal $\mathcal{P}_A$ from an aggregate release $\overline{A}^U$. We estimate the marginals via our denoising and debiasing techniques (from Section 4.3.1), depending on the application of privacy measures on $\overline{A}^U$.

Algorithm 3 describes our procedure for achieving an estimate $\hat{\mu}$ for the mean number of visits per user in the dataset given a privacy-aware aggregate release. Recall that $\mathcal{P}_A$ is set to $Exp(\hat{\mu})$. Algorithm 4 describes our procedure for computing which degree p will work best in the power transformation, to correct the empirical marginal $\widehat{\mathcal{P}}^0_-$ obtained directly from a ε-DP aggregate release.

Algorithm 1 GenerateSyntheticTrace

```
Inputs:
    $\mathcal{P}_S$: Approximated space marginal over ROIs
    $\mathcal{P}_T$: Approximated time marginal over epochs
    $\mathcal{P}_A$: Approximated activity marginal over trace sizes
    $DT(\mathcal{S})$: Delaunay triangulation of ROIs
Output:
    $L^s$: A synthetic trace
// We sample an origin ROI.
$s_0 \leftarrow$ sample_from_distribution($\mathcal{P}_S$, 1)
// Use $DT(\mathcal{S})$ to create a connected subgraph of ROIs including the origin ROI
$C(s_0) \leftarrow$ generate_connected_subgraph($s_0$, $DT(\mathcal{S})$, n_rois_subgraph = 10 (default value from [19])
// Normalize $\mathcal{P}_s$ restricted to $C(s_0)$.
$\mathcal{P}_{C(s_0)} \leftarrow$ normalize(restrict($\mathcal{P}_s$, $C(s_0)$))
// Sample the trace size (# visits).
n_visits ← round(sample_from_distribution($\mathcal{P}_A$, 1))
// Randomly sample n_visits ROIs and epochs with replacement.
ROIs ← sample_from_distribution($\mathcal{P}_{C(s_0)}$, n_visits)
epochs ← sample_from_distribution($\mathcal{P}_T$, n_visits)
return $L^s \leftarrow$ [ (ROIs[i], epochs[i]) for i = 1 ... n_visits ]
```

Algorithm 2 Approximate Marginals From Aggregate

```
Inputs:
    $\overline{A}^{\mathcal{U}}$ : Released aggregate
    $m$: Aggregate group size
    $p$: Specified probability distribution family
Output:
    $\mathcal{P}_S$: Approximated space marginal over ROIs
    $\mathcal{P}_T$: Approximated time marginal over epochs
    $\mathcal{P}_A$: Approximated activity marginal over trace sizes
// Compute direct estimates.
$\mathcal{P}_S, \mathcal{P}_T \leftarrow$ compute_empirical_marginals($\overline{A}^{\mathcal{U}}$)
$\mu^0_{visits} \leftarrow$ sum_entries($\overline{A}^{\mathcal{U}}$)$/m$
if $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}$ then
// Return direct estimates if no privacy.
    $\mathcal{P}_A \leftarrow$ fit_dist($p$, $\mu^0_{visits}$)
    return $\mathcal{P}_S, \mathcal{P}_T, \mathcal{P}_A$
end if
if $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}_{SSC}(k)$ and $k > 0$ then
// Apply log compression if SSC.
    $\mathcal{P}_S, \mathcal{P}_T \leftarrow$ log_compression($\mathcal{P}_S, \mathcal{P}_T$)
end if
if $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}_{DP}(\varepsilon)$ or $\overline{A}^{\mathcal{U}} = A^{\mathcal{U}}_{DP,SSC}(\varepsilon, k)$ then
// Apply power transformation if DP.
    $\mathcal{P}_S, \mathcal{P}_T \leftarrow$ power_transform($\mathcal{P}_S, \mathcal{P}_T$)
end if
// Apply Algorithm 3 from Appendix.
$\mu_{visits} \leftarrow$ estimate_mean($\mu^0_{visits}$, $A$, $m$, $\mathcal{P}_S, \mathcal{P}_T$, $k$, $\varepsilon$)
$\mathcal{P}_A \leftarrow$ fit_dist($p$, $\mu_{visits}$)
return $\mathcal{P}_S, \mathcal{P}_T, \mathcal{P}_A$
```

C ACCURACY RESULTS

In this section, we present the accuracy scores of ZK MIA and KK MIA for the experiments on suppression of small counts and ε-DP noise addition.

Table 3 presents the accuracy scores obtained by ZK and KK from the experiments on suppression of small counts from Section 6.1. Table 4 presents the accuracy scores obtained by ZK and KK from the experiments on event level ε-DP from Section 6.2. Table 5 presents the accuracy scores obtained by ZK and KK from the experiments on user-day level ε-DP.

We observe that the accuracy scores of KK and ZK are close in each experiment, as observed already with the AUC metric in the main text.

Algorithm 3 EstimateMean

```
Inputs:
    $\mu_0$: Initial guess for mean visits
    $A$ : Released aggregate
    $DT(\mathcal{S})$: Delaunay triangulation of ROIs
    $\mathcal{P}_s$: Estimated space marginal
    $\mathcal{P}_t$: Estimated time marginal
    $k$: Suppression threshold
    $\varepsilon, \Delta$: DP parameters
    $m$: Aggregate group size
Additional parameters
tol: Tolerance for stopping
max_iter: Max iterations
Output:
    $\mu$: Approximated mean visits per user
$\mu \leftarrow \mu_0$
for $i = 1$ to max_iter do
    $A_1 \leftarrow$ initialize_matrix()
// Create a synthetic aggregate of size $m$.
    for $j = 1$ to $m$ do
// Generate synthetic trace via with $\mu$ vistis.
        $A_1 \leftarrow A_1 +$ generate_synthetic_trace($\mathcal{P}_s, \mathcal{P}_t, \mu, DT(\mathcal{S})$)
    end for
    $A_1 \leftarrow$ apply_privacy_measures($A_1, k, \epsilon, \Delta$)
// Increase or decrease the estimate $\mu$ accordingly
    $\mu \leftarrow \mu_0 +$ (sum($A$)- sum($A_1$))$/m$
    if $|\mu - \mu_0| < tol$ then return $\mu$
    end if
end for
return $\mu$
```

Algorithm 4 pSelection

```
Inputs:
    $\sigma_0$: Reference variance
    $\mathcal{P}$: Space or time marginal to be modified
    $\epsilon_{tol}$: Error tolerance
Output:
    $p$: Degree for transformation $x^p$ that sets the variance of $\mathcal{P}$ to
    approximately match $\sigma_0$
$\sigma \leftarrow compute\_variance(\mathcal{P})$ // We compute the variance
from the original marginal value.
$p \leftarrow 1$
while $|\sigma_0 - \sigma| > \epsilon_{tol}$ do
    $\sigma \leftarrow compute\_variance(pow(\mathcal{P}, p))$
// Increment the power $p$ until estimate is in range.
    $p \leftarrow p + 0.01$
end while
return $\sigma$
```

Table 3: Mean accuracy scores with standard error for KK and ZK on size 1000 aggregates from the CDR and Milan datasets across various suppression thresholds k.

k	CDR dataset		Milan dataset	
	KK	ZK	KK	ZK
0	0.980 ± 0.012	*0.991 ± 0.008*	0.990 ± 0.005	*0.990 ± 0.003*
1	*0.907 ± 0.028*	0.879 ± 0.025	*0.767 ± 0.018*	*0.700 ± 0.017*
2	0.807 ± 0.031	*0.827 ± 0.026*	*0.683 ± 0.019*	*0.631 ± 0.013*
3	*0.685 ± 0.032*	0.687 ± 0.031	*0.600 ± 0.016*	*0.550 ± 0.009*
4	*0.597 ± 0.024*	0.603 ± 0.027	*0.566 ± 0.011*	*0.512 ± 0.003*
5	0.543 ± 0.018	*0.528 ± 0.019*	*0.536 ± 0.010*	*0.500 ± 0.000*

Table 4: Mean accuracy scores with standard error for KK and ZK on size 1000 aggregates from the CDR and Milan datasets across various privacy budgets ε for event level DP.

ε	CDR dataset		Milan dataset	
	KK	ZK	KK	ZK
0.1	*0.588 ± 0.019*	0.555 ± 0.014	*0.539 ± 0.006*	*0.549 ± 0.007*
0.5	*0.848 ± 0.028*	0.791 ± 0.019	*0.744 ± 0.010*	*0.634 ± 0.007*
1.0	0.920 ± 0.026	0.907 ± 0.018	*0.850 ± 0.016*	*0.594 ± 0.008*
5.0	0.906 ± 0.035	0.934 ± 0.019	*0.881 ± 0.022*	*0.660 ± 0.018*
10.0	0.923 ± 0.029	0.934 ± 0.018	*0.920 ± 0.021*	*0.671 ± 0.019*

Table 5: Mean accuracy scores with standard error for KK and ZK on size 1000 aggregates from the CDR and Milan datasets across various privacy budgets ε for user-day level DP.

ε	CDR dataset		Milan dataset	
	KK	ZK	KK	ZK
0.1	0.502 ± 0.014	*0.502 ± 0.010*	0.497 ± 0.006	0.496 ± 0.006
0.5	0.508 ± 0.014	*0.526 ± 0.016*	0.517 ± 0.006	0.519 ± 0.006
1.0	0.533 ± 0.014	*0.539 ± 0.014*	0.534 ± 0.006	0.544 ± 0.007
5.0	*0.723 ± 0.020*	0.676 ± 0.018	0.746 ± 0.009	0.680 ± 0.008
10.0	*0.874 ± 0.025*	0.825 ± 0.019	0.870 ± 0.014	0.777 ± 0.018

D ADDITIONAL EXPERIMENTS

D.1 Varying the size of the aggregate

Since ZK MIA requires the estimation of statistics from the aggregate, there may be concerns about its performance when the aggregate size is small. However, like previous MIAs, ZK MIA performs more effectively on smaller-scale aggregates compared to larger aggregates. This is shown in Figure 11 for aggregate sizes $m = 100, 250, 500, 1000$ and different privacy budgets ε.

To further understand how MIA performance scales with aggregate size m, we also consider $m > 1000$ in this experiment. To this end, we vary $m = 100, 500, 1000, 2000, 3000$ and compare the performance of KK MIA and ZK MIA on raw ($k = 0$) and suppressed ($k = 1$) aggregates. Results on the CDR dataset are reported in Tables 6 and 8 and results on the Milan dataset are reported in Tables 7 and 9. $m = 3000$ was not run on the Milan dataset due size limitations.

In these settings with mild privacy protection, the attacks always succeed regardless of the value of m. We also observe a few intuitive trends. First, when raw aggregates ($k = 0$) are attacked, increasing the size of the aggregates slowly decreases the performance of the attack. On the CDR dataset, KK and ZK attain AUCs 0.999 and 1.0 for $m = 100$, which decreases to 0.919 and 0.977 for $m = 3000$. Second, when we apply suppression $k = 1$, the attacks initially perform poorly when the aggregate size is small. We hypothesize this to be due to a larger percentage of entries being suppressed when fewer traces are aggregated, leaving less information in the release. This effect gradually decrease as aggregate size increases. It is then counterbalanced by the first effect, that increasing the size of the aggregates slowly decreases the performance of the attack, when aggregate sizes increase. This is visible for $m \leq 1000$ in the CDR dataset. For the Milan dataset, AUC however still monotonically increases even beyond $m \leq 1000$ as the dataset is more sensitive to suppression with the average user has approximately 6 times less visits, as shown in Table 14b.

Table 6: Mean AUCs of KK and ZK MIAs for $k = 0$ on the CDR dataset with varying m.

m	KK	ZK
100	1.000 ± 0.000	1.000 ± 0.000
500	1.000 ± 0.000	1.000 ± 0.000
1000	1.000 ± 0.000	0.999 ± 0.001
2000	0.997 ± 0.003	0.994 ± 0.006
3000	0.988 ± 0.011	0.977 ± 0.021

Table 7: Mean AUCs of KK and ZK MIAs for $k = 0$ on the Milan dataset with varying m.

m	KK	ZK
100	1.000 ± 0.000	1.000 ± 0.000
500	1.000 ± 0.000	1.000 ± 0.000
1000	0.995 ± 0.002	1.000 ± 0.000
2000	0.977 ± 0.011	1.000 ± 0.000

Table 8: Mean AUCs of KK and ZK MIAs for $k = 1$ on the CDR dataset with varying m.

m	KK	ZK
100	0.856 ± 0.019	0.779 ± 0.041
500	0.961 ± 0.008	0.987 ± 0.003
1000	0.981 ± 0.007	0.976 ± 0.010
2000	0.979 ± 0.010	0.965 ± 0.013
3000	0.973 ± 0.011	0.938 ± 0.016

D.2 Increasing the size of ZK synthetic reference

Figure 12 illustrates how increasing the number of synthetic traces available to the attacker improves the MIA's performance up to marginal returns.

Table 9: Mean AUCs of KK and ZK MIAs for $k = 1$ on the Milan dataset with varying m.

m	KK	ZK
100	0.756 ± 0.020	0.701 ± 0.046
500	0.889 ± 0.018	0.885 ± 0.025
1000	0.916 ± 0.015	0.919 ± 0.016
2000	0.981 ± 0.009	0.972 ± 0.007

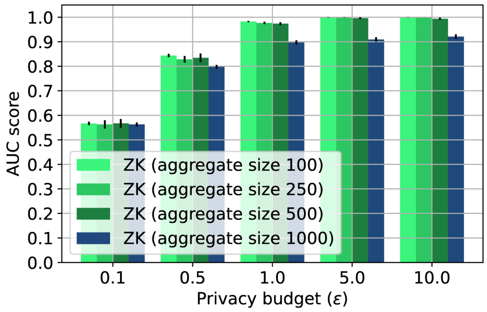


Figure 11: Mean AUC scores with standard error for ZK on event level ε-DP for varying privacy budgets ε on aggregates of varying sizes from the Milan dataset.

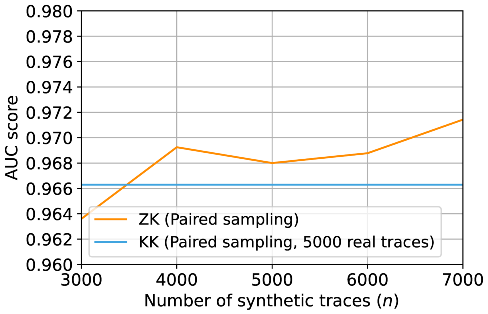


Figure 12: AUC of the Zero Auxiliary Knowledge MIA for different number of synthetic traces generated in the setting $m = 1000$, $k = 1$, and $\varepsilon = 1$-DP at the event level.

D.3 No time information

In this experiment, we now assume that the adversary only has access to some of the locations that the target has visited, without knowing the epochs during which the visits were done. For example, the adversary may know the target's home and work. To model this attack setting, we suppose that the adversary either knows the target's top-K most visited ROIs, for $K = 1, 2, 3$, or the full set of the target's visited ROIs during the observation period. In one implementation, which we call "greedy", the adversary assumes that the target visits each known ROI during every epoch in the observation period. This ensures that the visits to these ROIs are reflected in the target trace, but it also sets many incorrect visits.

Results are presented in Table 11. In our second implementation, which we call "random sampling", the adversary distributes the target's visits uniformly across the known ROIs. For example, if the adversary knows the top-3 ROIs, and their estimate for the mean number of visits per user is μ, then they would sample $\frac{\mu}{3}$ visits for each of the top-3 ROIs. The corresponding epochs for each visit are sampled from the estimated time marginal. For simplicity, we assume that μ is the true mean number of visits and that the estimated time marginal is the true one. Results on raw aggregates of size $m = 1000$ are presented in Table 10.

Table 10 shows that both MIAs perform poorly ($AUC < 0.63$) when the adversary uses random sampling to approximate the target trace. This suggests that random sampling fails to estimate the target trace, due to the omission of true target visits, and the inclusion of incorrect visits.

In contrast, Table 11 shows that KK was able to perform significantly better than random when the adversary knew more than 2 of the target's most visited ROIs and used the greedy implementation (ex. $AUC = 0.86$ on Milan when knowing all visited ROIs). This suggests that, although the greedy implementation includes many incorrect visits, the guaranteed inclusion of some of the target's actual visits enables membership inference to an extent. ZK, on the other hand, fails to attain 0.6 AUC. Since ZK already replaces real individual traces with synthetic traces, we hypothesize that membership inference becomes too difficult if the estimated target trace contains significantly incorrect information.

We however note that our current implementation for sampling the visits under this prior knowledge might be suboptimal and that better implementations might exist. For example, [78] uses a synthetic target trace, using social network information and the traces of the target's friends. We leave this exploration for future work.

E ADDITIONAL PLOTS

We present additional figures demonstrating statistics related to the location datasets.

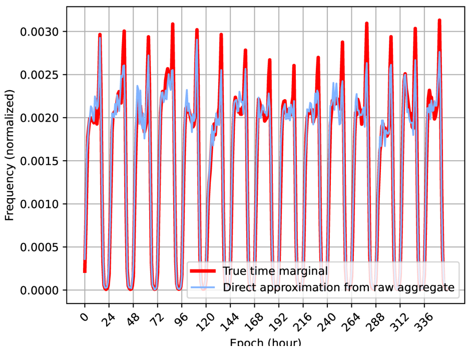
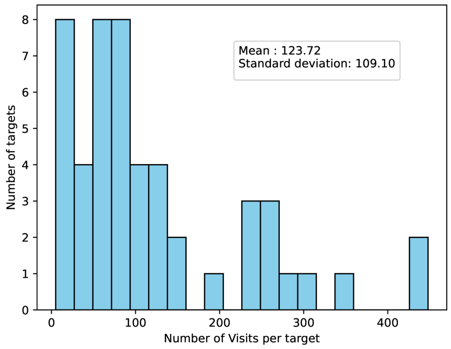
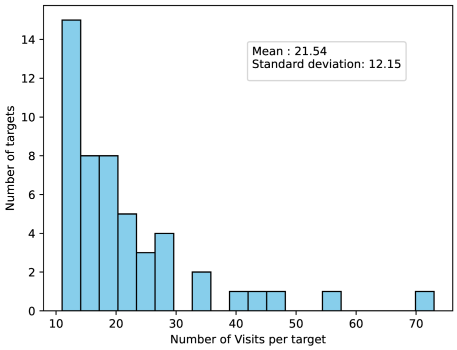


Figure 13: Time marginal from a raw aggregate over $m = 1000$ users from the CDR dataset.



(a) CDR dataset



(b) Milan

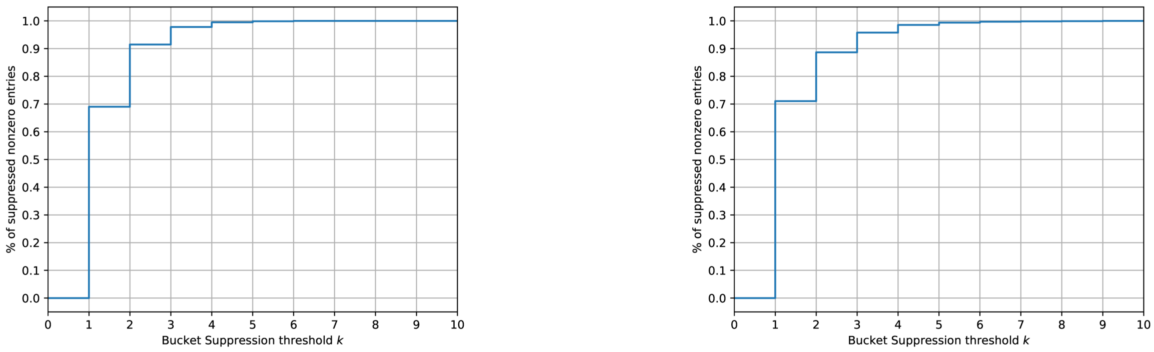
Figure 14: Number of visits per target over the 50 targets.

Dataset	Knock Knock				Zero Auxiliary Knowledge			
	Top 1	Top 2	Top 3	All	Top 1	Top 2	Top 3	All
CDR	0.542 ± 0.04	0.538 ± 0.03	0.531 ± 0.02	0.525 ± 0.02	0.524 ± 0.03	0.528 ± 0.02	0.515 ± 0.02	0.510 ± 0.02
Milan	0.576 ± 0.02	0.628 ± 0.03	0.614 ± 0.03	0.560 ± 0.03	0.518 ± 0.02	0.553 ± 0.02	0.568 ± 0.02	0.556 ± 0.04

Table 10: Mean AUC scores with standard error for KK and ZK on raw aggregates of size $m = 1000$ when the adversary only knows some of the target's visited ROIs and employs the random sampling approach of distributing random visits uniformly across each known ROI.

Dataset	Knock Knock				Zero Auxiliary Knowledge			
	Top 1	Top 2	Top 3	All	Top 1	Top 2	Top 3	All
CDR	0.571 ± 0.05	0.611 ± 0.05	0.647 ± 0.06	0.825 ± 0.05	0.512 ± 0.02	0.527 ± 0.01	0.514 ± 0.02	0.516 ± 0.01
Milan	0.682 ± 0.03	0.764 ± 0.03	0.822 ± 0.03	0.860 ± 0.02	0.542 ± 0.02	0.524 ± 0.02	0.542 ± 0.02	0.545 ± 0.02

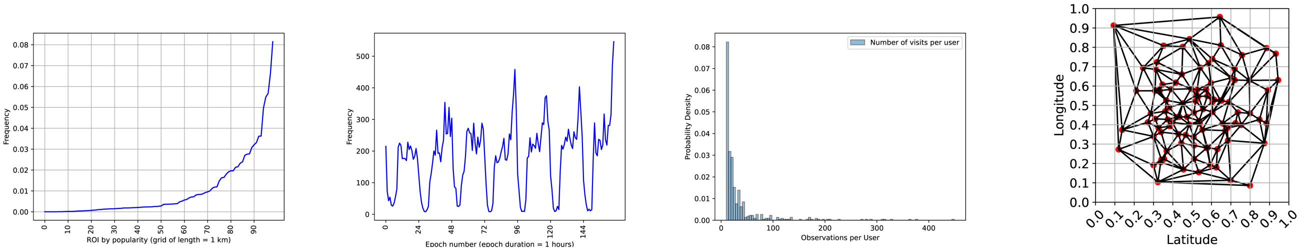
Table 11: Mean AUC scores with standard error for KK and ZK on raw aggregates of size $m = 1000$ when the adversary only knows some of the target's visited ROIs and employs the greedy approach of assuming that the target visits each known ROI during every epoch.



(a) CDR dataset

(b) Milan

Figure 15: The percentage of nonzero entries that are suppressed in a size $m = 1000$ aggregate after undergoing SSC with threshold k is plotted.



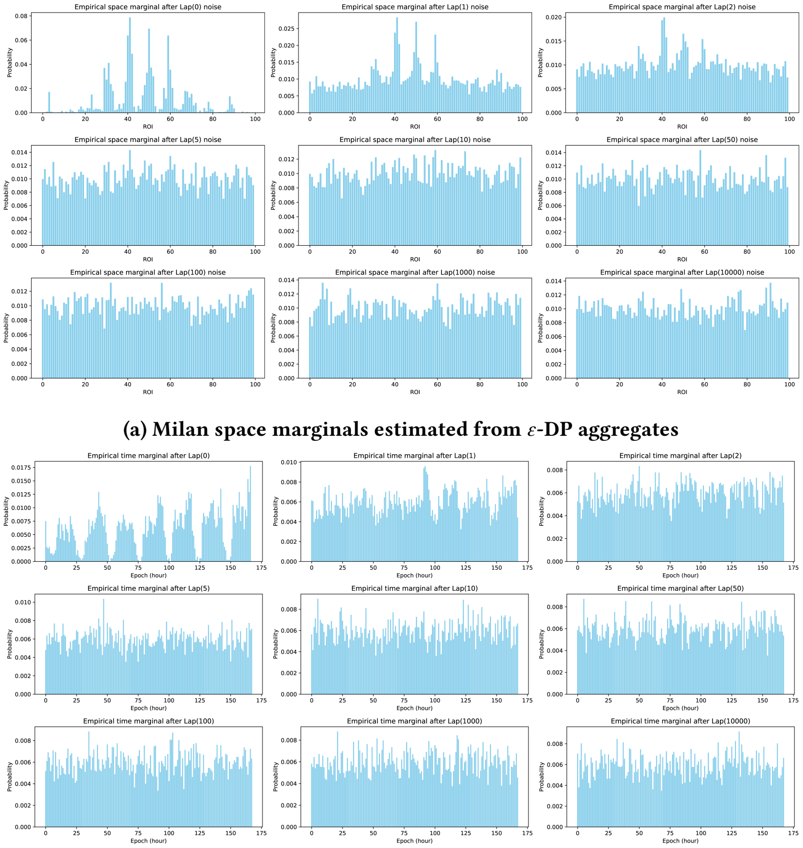
(a) Space Marginal

(b) Time Marginal

(c) Activity Marginal

(d) Delaunay Triangulation

Figure 16: The four statistical parameters for the unicity model by Farzanehfar et al. [19] include marginal distributions in Figures 16a-16c (the dataset's true marginal distributions are shown in red) and the Delaunay triangulation of ROIs in Figure 16d.



(a) Milan space marginals estimated from ε-DP aggregates

(b) Milan time marginals estimated from ε-DP aggregates

Figure 17: The space and time marginals directly obtained from ε-DP aggregates over $m = 1000$ users from the Milan dataset are plotted for different noise scales $\frac{\Delta}{\varepsilon}$. Interestingly, the distribution does not converge to a uniform distribution as the noise scale increases, due to the increasing variance of $Lap(\frac{\Delta}{\varepsilon})$