Can Knowledge of Demographics and Privacy Parameters Break Location Privacy?

Maja Schneider[1][a], Charini Nanayakkara[2][b], Peter Christen[2][c], Erik Buchmann[1][d] and Erhard Rahm[1][e]

[1]*Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany*

[2]*School of Computing, The Australian National University, Canberra, Australia*

{mschneider, buchmann, rahm}@informatik.uni-leipzig.de, {charini.nanayakkara,peter.christen}@anu.edu.au

Keywords: CONTEXT LINKING ATTACK, DIFFERENTIAL PRIVACY, SEMANTICS, POINTS OF INTEREST

Abstract: Location-based applications offer increasingly personalized services to mobile users. Incorporating temporal and demographic information can further improve service quality. However, sharing such information carries the risk of leaking private data, including a user's identity or further personal attributes. Differential Privacy (DP) is a widely accepted privacy notion to protect user data in this context. However, DP does not account for adversarial background knowledge, which can undermine privacy through context linking attacks. To design resilient privacy mechanisms, a systematic analysis is required to determine which pieces of background information pose the highest risk. In this work, we investigate whether knowing the privacy mechanism and semantic information can break DP and enable an adversary to reconstruct a user's location. We evaluate which types of background knowledge contribute most to attack success by designing a series of attacks with increasing access to semantic context, such as points of interest (POIs), mobility statistics, demographic data, and privacy parameters. We conduct an extensive evaluation on two large datasets. Our results show that knowledge of POIs and typical mobility patterns, especially when combined with the privacy parameter, substantially increases attack success, particularly in rural areas and for certain demographic groups.

1 INTRODUCTION

With the continuous advancement of mobile applications, users increasingly benefit from personalized and context-aware location-based services (LBS). While the user's spatiotemporal information is key to deliver core functionality, service quality can be enhanced by incorporating sociodemographic information about the user. For instance, an LBS recommending nearby restaurants to a user might propose inexpensive dining options to a student, but higher-quality restaurants to a full-time professional.

However, sharing such information may expose private details about users. From spatiotemporal data, adversaries can infer potentially sensitive points of interest (POIs), such as a user's home or workplace, medical facilities or religious institutions, and may thus derive personal attributes or even re-identify individuals (Liao et al., 2005; Hoh et al., 2006; Krumm, 2007).

Privacy-preserving mechanisms are therefore crucial to protect users in such a setting. However, they are increasingly challenged by adversaries with substantial background knowledge who perform context linking attacks (Wernke et al., 2014) allowing them to reconstruct the user's true location. Differential Privacy (DP) (Dwork et al., 2006) is a widely adopted privacy notion in this setting, but it does not explicitly model adversarial background knowledge, leaving DP-based mechanisms vulnerable when attackers exploit semantic data or user-specific priors.

Prior work mainly considers generic background knowledge derived from population-level mobility statistics or POI data (Chatzikokolakis et al., 2015; Li et al., 2018; Tian et al., 2021). However, the impact of user-specific knowledge, particularly sociodemographic information, remains largely unexplored. Furthermore, there is little guidance on how privacy parameters should be chosen when attackers possess different types and levels of background knowledge.

[a] https://orcid.org/0000-0001-5936-1415
[b] https://orcid.org/0000-0002-7603-1845
[c] https://orcid.org/0000-0003-3435-2015
[d] https://orcid.org/0009-0009-5874-4313
[e] https://orcid.org/0000-0002-2665-1114

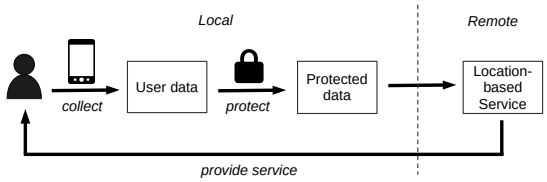


Figure 1: The system architecture.

This work addresses these gaps by systematically analyzing which forms of background knowledge most affect the success of context linking attacks, and by providing a detailed evaluation of how sociodemographic information influences these attacks.

We consider the system architecture, shown in Figure 1, and illustrate the problem with the following use case: *A mobile user likes to obtain a personalized service from an app on their phone, for example to find a restaurant close by that suits their taste and is currently open. The app first requests information from the user about their location, the current time and some demographic information to tailor its results. Concerned about their privacy, the user locally protects their data by, for example, applying a DP-based mechanism, before sharing their data. The app itself might offer such protection, where it sends the protected data to an LBS which calculates and returns a personalized recommendation. With increasing level of privacy protection, the result becomes less accurate and relevant for the user, while weaker protection increases the risk of privacy leakage. The user therefore faces the challenge of understanding which pieces of information, whether stemming from personal data, external semantic information, or from the employed privacy mechanism, pose the highest risk and how to configure their privacy parameters to balance utility and privacy.*

Against this background, we address the following research questions: *(1) How well can an adversary reconstruct a user's true location from a protected one by exploiting background knowledge about the user? (2) Which types of background information carry the highest risk of successful reconstruction?*

We study these questions through a systematic empirical analysis of context linking attacks with increasing adversarial knowledge. In our setting, the sensitive information comprises the user's true location and attributes that can be inferred from it, such as the type of place visited. An adversary attempts to recover this information by defining a circular attack region around the user's observed location, assigning probabilities to each point within it, and selecting the location (or POI) with the highest likelihood as the predicted true location. We examine how this reconstruction ability changes when the adversary has access to additional background knowledge, specifically, details about the employed privacy mechanism and its parameters (such as the privacy budget), as well as semantic knowledge like the distribution of POIs, mobility patterns, and sociodemographic data.

Our specific contributions are:

- We formalize the attack scenario and describe the involved building blocks: the user, the privacy mechanism, the adversary with their background knowledge, and the attack approach.
- We design several attacks, that gradually increase the background knowledge of the adversary. In particular, we propose new semantic attacks that exploit user-specific knowledge, such as a user's sociodemographic features.
- We evaluate our research questions using two large data sets, one synthetic and real-world. Our experiments show that semantic knowledge carries a high risk of leaking a user's true location. The risk is particularly high when semantic knowledge is combined with knowledge of the privacy mechanism and its parameters.

The rest of the paper is structured as follows. Section 2 reviews related work and background. Section 3 introduces the problem, followed by an evaluation and its results in Section 4 and 5. The approach is discussed in Section 6. Section 7 concludes.

2 RELATED WORK AND BACKGROUND

In this section, we review related work with regards to context linking attacks and defenses.

2.1 Context Linking Attacks

In a *localization attack*, an adversary tries to infer a user's true location x from an observation, and might even try to re-identify the user with an *identity attack*, for example by linking the user's home location and identity (Liu et al., 2018). The adversary's observation can be a single snapshot location or a historical trace, gathered by repeated measurements of a user's position over time. In our work, we focus on snapshot locations and thus describe related work from this context.

One approach to realize a localization attack is the *context linking attack* (Wernke et al., 2014). In this attack, an adversary uses contextual background knowledge to increase their certainty about the user's true location and potentially learn further information.

Such background information can be anything that is helpful to understand the user's mobility pattern and to limit the user's possible places of presence. This can be publicly available data, such as distributions of users, traffic or travel statistics, or map information, for example POI distributions, road networks, or an opening hours directory (Shokri et al., 2011; Liu et al., 2018). With such knowledge, an adversary can exclude rarely frequented or unlikely regions, such as lakes, deserts, or closed shops. From statistical mobility data, the adversary can create a probability distribution indicating the likelihood of a user to be at a certain location (Hoh et al., 2006; Krumm, 2007; Isaacman et al., 2011). Personal information about a user, such as a their sociodemographic characteristics, social relationships or POI preferences, can be useful to further refine the probability distribution (Sadilek et al., 2012).

(Tian et al., 2021) propose the semantic-relativity attack which exploits background knowledge of location semantics of the road network and an individual sensitivity setting regarding certain location semantic types by the user. Other works consider social media content, such as tweets to infer a user's home location (Mahmud et al., 2014). (Cao et al., 2018) show that an adversary can exploit a city's unique spatial structure, such as street layouts and POI distributions, to reconstruct a user's location when the user shares their nearby POIs. While related work shows that sociodemographic information can be revealed from location data (Zhong et al., 2015; Li et al., 2018; Wu et al., 2019), research is missing on the impact of such personal information on recovering a user's location.

2.2 Differential Privacy

A widely used standard for privacy protection is Differential Privacy (DP) (Dwork et al., 2006). DP provides a formal privacy guarantee by ensuring that the output of a query over a database does not significantly change when a single individual's data is added to or removed from the database. The level of privacy is controlled by a privacy budget ε. DP can be achieved by perturbing the data with a noise function, parameterized by ε. Originally designed for protecting an aggregation result from a database query, the concept was extended to publish differentially private static data sets later on (Guerra-Balboa et al., 2022).

DP has been adapted for the case of publishing location data, called *geo-indistinguishability* (Andrés et al., 2013). In this context, DP ensures that two locations are indistinguishable from each other within a radius r. This means that the distance between the distributions $K(x)$ and $K(x')$, produced by the locations x and x', is at most $\varepsilon \cdot r$ for all locations x' with $d(x,x') \leq r$ where $d(\cdot,\cdot)$ is the Euclidean distance. Formally, a mechanism K satisfies ε-geo-indistinguishability iff for all x, x':

$$d_{\mathcal{P}}(K(x),K(x')) \leq \varepsilon \cdot d(x,x') \tag{1}$$

where $d_{\mathcal{P}}(\cdot,\cdot)$ is the multiplicative distance between two distributions σ_1, σ_2, defined on a set $\mathcal{S}$ as

$$d_{\mathcal{P}}(\sigma_1,\sigma_2) = \sup_{S \subseteq \mathcal{S}} \left| \ln \frac{\sigma_1(S)}{\sigma_2(S)} \right|. \tag{2}$$

To instantiate geo-indistinguishability in the continuous plane, the planar *Laplace mechanism* can be used (Dwork, 2011; Andrés et al., 2013). Instead of reporting the true location x, an obfuscated location x' is produced by adding random noise. Noise is sampled from the Laplace function, centered at x, with a probability density function:

$$pdf(x') = \frac{\varepsilon}{2} e^{-\varepsilon|x'-x|}. \tag{3}$$

2.3 Privacy Mechanisms using Semantic Background Knowledge

Several studies indicate that DP can still be compromised when semantic background information is taken into account. (Primault et al., 2014) show that a large number of locations can be reconstructed when an adversary is linking obfuscated locations with POI distributions. Going further, (Maouche et al., 2017) show that re-identification even is possible in many cases if such distributions are known. Therefore, privacy protection mechanisms need to integrate background knowledge in the protection scheme as well, and make sure that users understand how to set the privacy parameters (Schneider et al., 2024).

In particular, these mechanisms need to ensure that the semantic properties of location data are protected (Guerra-Balboa et al., 2022). For example, when using DP without considering semantics, locations might be moved to unreachable areas, such as water bodies. Furthermore, the semantic information could remain present in the data and thus potentially be identified. Some location privacy-preserving mechanisms, such as dummy or cloaking approaches, aim to create reasonable obfuscations and obtain semantic diversity in the result set in order to protect a user from context linking attacks (Shen et al., 2020; Zhang et al., 2022).

Several extensions to geo-indistinguishability have been proposed to defend against adversarial background knowledge. (Yan et al., 2022) extend the Laplace mechanism by an additional step that ensures that obfuscated locations possess sufficiently

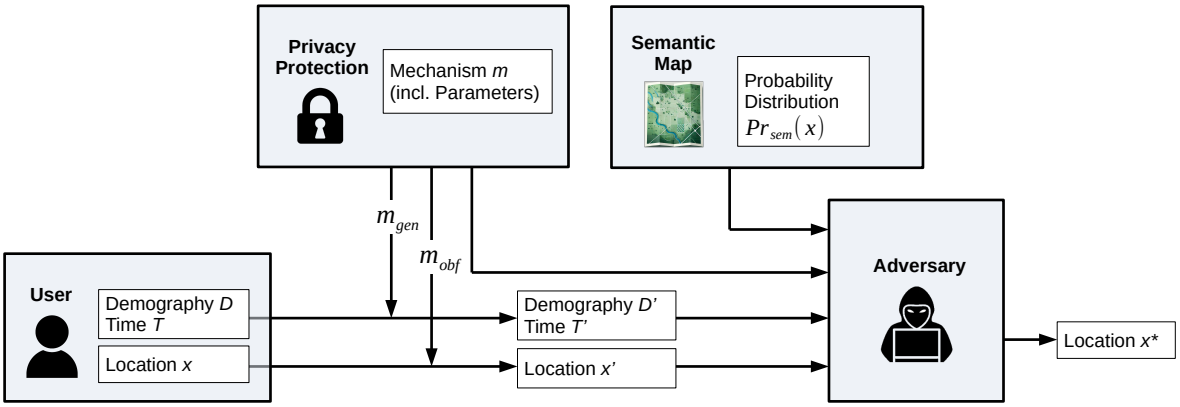


Figure 2: The attack model.

large user populations and dissimilar POI semantics. Similarly, (Chatzikokolakis et al., 2015) enhance geo-indistinguishability through an adaptive component that adjusts the geometrical obfuscation distance based on the privacy mass, a concept that captures spatial and semantic features, such as population density and POIs. (Li et al., 2018) demonstrate that shared check-ins enable semantic profiling of users and mitigate this risk with a POI- and activity-aware obfuscation method. (Peng et al., 2019) take into account location correlation between multiple users to create a personalized privacy mechanism that enhances the traditional DP-mechanism.

3 PROBLEM DESCRIPTION

Following (Shokri et al., 2011), we formalize the four building blocks of the context linking attack: (1) the user and their data, (2) the privacy protection, (3) the background knowledge of the adversary, and (4) the adversary's attack approach. The interplay of these elements is illustrated in the attack model in Figure 2.

A user u shares their data with a service provider, which acts as a curious adversary. This means, they use external knowledge to find out as much as possible about the user. The user's data is obfuscated with a privacy protection mechanism. The adversary has certain background knowledge and uses this to predict the true location of u, given the observed, protected location. To carry out this inference, the adversary defines a circular region centered on the observed location, assigns probabilities to each location within this area, and selects the POI with the highest estimated likelihood. Considering this scenario, our goal is to analyze if the adversary can successfully reconstruct the user's location and which part of the background knowledge contributes the most to a successful attack.

3.1 User

We consider a user $u \in U$ with a set of certain sociodemographic attributes D. The user u moves in a geographical region R and shares their current whereabouts with a service provider in order to retrieve a LBS, such as to find the closest restaurant to u's current location. The shared information includes the user's current location $x_u(t) \in R$ and timestamp t. The location x_u is a spatial point, for example described as a pair of latitude and longitude. From t, a set of temporal attributes T can be derived, containing (besides to the timestamp), for example, the day of the week.

R contains a set of POIs P. A POI $p \in P$ is described as a tuple $p = (x_p, o_p, l_p)$. o_p is the outline of p and the location x_p is its centroid. The label l_p is a semantic description of p, for example, a *park* or an *apartment*. At the time t of sending a request to the LBS, we assume the user u to be located at a POI p, so that $x_u(t) \in o_p$. At this POI, u performs a certain activity $a_u(t) \in A$, for example, *relaxing* or *sleeping*.

3.2 Privacy Protection

On the location x_u, the user applies a privacy protection mechanism m_{obf} using certain privacy parameters. We assume m_{obf} to be the Laplace mechanism (Andrés et al., 2013), as described in Section 2.2. This mechanism produces an obfuscated location x' from the true location x by adding noise from the Laplace distribution.

Furthermore, the user protects their sociodemographic attributes D and the temporal attributes T using a generalization approach m_{gen}, which reduces the precision of the attributes in D and T to obtain protected versions D' and T' (Samarati and Sweeney, 1998). Increasing levels of generalization g reduce the granularity of D and T.

3.3 Background Knowledge of the Adversary

The adversary tries to infer the user's true location x after observing the protected location x'. The adversary has certain background information about the user u, which can include u's protected sociodemographic and temporal attributes D' and T', the privacy protection mechanisms m_{obf} and m_{gen} that u used, and the employed privacy parameters. In many real-world services, details about the applied privacy mechanism and its parameters are disclosed to build trust, meaning that the adversary may also access this information. The adversary might further have access to a population-wide mobility data set that includes sociodemographic attributes, timestamps, and activities of persons, which can also overlap with the users U. Furthermore, the adversary can access a service to query all POIs $p \in P$ in the region R with their label and GPS coordinates.

We consider two types of background knowledge that an adversary can have, namely *privacy knowledge* (**Lap** and **Eps**), and *semantic knowledge* (**POI**, **Mob**, and **Dem**):

Lap: The adversary knows that the user used the Laplace mechanism to obfuscate their location. From this information, the adversary can guess the noise function and better estimate which regions are more likely to contain the user's obfuscated location.

Eps: The adversary knows the privacy budget ε, used to obfuscate the user's location, allowing them to better estimate the obfuscation distance. Such information can be obtained, for example, when an app with hard-coded privacy parameters is used.

POI: The adversary has knowledge of the distribution of POIs in R. This information significantly limits the potential places where the user can be.

Mob: The adversary has access to a population-wide mobility data set of persons, indicating the POIs they visited during a day, from which the adversary can derive a generic mobility profile. We hypothesize that such information helps the adversary to identify regions that are more likely for the user to be in.

Dem: The adversary has access to a set of sociodemographic attributes of each person in the mobility data set in **Mob**, allowing them to derive a specific mobility profile for a user with such attributes. For example, if an adversary knows that a user is a young student, they may assign a higher likelihood to locations at educational institutions, such as schools.

Table 1: Attacks and their background knowledge.

Attack	Privacy Knowledge	Semantic Knowledge
N	-	-
L	Lap	-
LE	Lap, Eps	-
P	-	POI
PM	-	POI, Mob
PMD	-	POI, Mob, Dem
LE-P	Lap, Eps	POI
LE-PM	Lap, Eps	POI, Mob
LE-PMD	Lap, Eps	POI, Mob, Dem

3.4 Attacks

Depending on the background knowledge of the adversary described in Section 3.3, we define different attack strategies that require either privacy knowledge, semantic knowledge, or both. As a baseline, the adversary has no knowledge at all. The nine different attacks and their respective background knowledge are summarized in Table 1 and described in detail in the following subsections.

3.4.1 Attacks with Privacy Knowledge

In each attack, the basic strategy is the same. The adversary creates a circular region $C_{attack} \subset R$ centered at the observed location x' with radius r_{attack}. Each location $x \subset C_{attack}$ is weighted with a probability distribution $Pr(\cdot)$, indicating the likelihood of presence of the user. The adversary makes a prediction x^* of the user's true location by calculating the location with the highest likelihood:

$$x^* = \underset{x}{\text{argmax}}\ Pr(x), x \subset C_{attack}. \tag{4}$$

Figure 3 shows an example of such an attack on a student visiting a school.

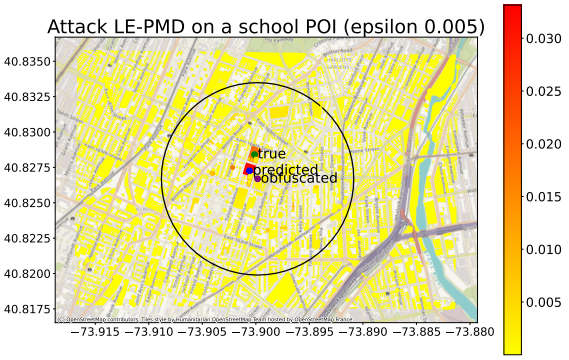


Figure 3: Attack on the obfuscated location ($\varepsilon = 0.005$) of a student visiting a school. The probability distribution is visualized by color. The adversary selects the most likely location (red region) within the attack circle C_{attack}.

Attack *N*: Given no background knowledge at all, the adversary randomly chooses $r_{attack} \in (0, r_{max}]$, where r_{max} is the maximum attack radius and a parameter of the attack. The adversary creates $Pr(\cdot)$ from an even distribution over all $x \subset C_{attack}$, making each location equally likely to be selected.

Attack *L*: Given **Lap**, the adversary randomly chooses an $\varepsilon \in (0, \varepsilon_{max}]$, where ε_{max} is a parameter of the attack. Based on ε, the adversary estimates the obfuscation distance to determine the attack radius r_{attack}. For each candidate location $x \subset C_{attack}$ at distance $r = d(x, x')$ from the observed location x', the adversary calculates $Pr(\cdot)$ as the planar Laplace likelihood $L(x) = e^{-\varepsilon r}$. This increases the likelihood for the adversary to pick locations close to x'.

Attack *LE*: Given **Lap** and **Eps**, the adversary uses the true value of ε and proceeds otherwise as described in attack ***L*** above.

3.4.2 Attacks with Semantic Knowledge

In this type of attack, the adversary uses semantic information about POIs and user profiles from mobility training data sets to create a more refined probability distribution $Pr(\cdot)$. Using a public service, the adversary obtains all POIs $p \in P$ in the region R and constructs a *semantic map*. This map indicates, for each location x, whether it is associated with a POI, and assigns a likelihood $Pr_{sem}(x)$ for the user u to be at x.

Attack *P*: The adversary proceeds as described in attack ***N*** to construct the attack circle. Given **POI**, the adversary then calculates $Pr(\cdot)$ as a probability distribution $Pr_{sem}(\cdot)$ that allows them to pick any POI p in the attack circle with an even likelihood:

$$Pr_{sem}(x) = \begin{cases} \frac{1}{|\{p : p \subset C_{attack}\}|}, & \text{if } is\text{-}poi(x) \\ 0, & \text{else,} \end{cases} \quad (5)$$

where $is\text{-}poi(x) = \exists p \subset P : x = x_p$.

Attack *PM*: Given **POI** and **Mob**, the adversary uses a mobility training data set to calculate a generic mobility profile, which reflects the movement behavior of all users in the data set. From this profile, the adversary constructs the probability distribution $Pr_{sem}(x)$ proportionally to the frequency of presence at location x. The adversary runs all other steps as in attack ***P***, but refines $Pr_{sem}(\cdot)$ as described.

Attack *PD*: Given **POI** and **Dem**, the adversary's goal is to calculate a specific mobility profile for the user u. To this end, the adversary selects from the mobility training data set all persons that have the same (protected) background information D' and T' as user u and creates the probability distribution $Pr_{sem}(\cdot)$ as described in attack ***PM*** and proceeds otherwise also as in attack ***PM***.

3.4.3 Attacks with Combined Knowledge

Attacks ***LE-P***, ***LE-PM***, and ***LE-PMD*** have full privacy knowledge and differing semantic knowledge. The adversary thus creates a semantic map according to the given semantic knowledge, and then applies a weighting with the planar Laplace likelihood as described in attack ***LE*** above.

Attack *LE-P*: Given **Lap**, **Eps**, and **POI** the adversary runs all steps from attack ***P***, but constructs $Pr(\cdot)$ by weighting $Pr_{sem}(\cdot)$ with the planar Laplace likelihood as described in attack ***LE***.

Attack *LE-PM*: Given **Lap**, **Eps**, **POI**, and **Mob**, the adversary runs all steps from attack ***PM***, but constructs $Pr(\cdot)$ by weighting $Pr_{sem}(\cdot)$ with the planar Laplace likelihood as described in attack ***LE***.

Attack *LE-PMD*: Given **Lap**, **Eps**, **POI**, **Mob**, and **Dem**, the adversary runs all steps from attack ***PMD***, but constructs $Pr(\cdot)$ by weighting $Pr_{sem}(\cdot)$ with the planar Laplace likelihood as described in attack ***LE*** above.

4 EXPERIMENTS

We now evaluate an adversary's attack success using the nine different attacks described above.

4.1 Data Sets

For our evaluation we use two large population-wide mobility data sets that can be linked to sociodemographic and temporal features of the data producers.

ASTRA: We create synthetic mobility data with the ASTRA data generation tool (Schneider et al., 2026) in the region of New York City, USA for 10,000 artificial users. This approach creates POI trajectories by translating travel diaries of real persons into geographical traces. We use the Multinational Time Use Study (MTUS) (Fisher et al., 2022; Gershuny et al., 2020) data set for the USA between January 2003 and December 2019 as the travel diary data basis. Each POI trajectory comprises a user's daily history of POIs along with each POI's GPS location, consisting of latitude and longitude, and the timestamp of the visit. Based on the timestamp, we derive the season and the weekday of the visit.

Within this data set, the user's sociodemographic features are known. These include basic demographic features (*age, sex,* and *citizenship*), the user's educational and financial background (*education level, employment status, working time, income, retired,* and *student*), details about the user's private life (*number*

of children, age of youngest child, and *single parenting*), and information about the user's living situation (*urban living,* and *house ownership*).

The age and sex of the synthesized users corresponds to the true distribution of these attributes in the selected simulation region, obtained from census data. Their sociodemographic background is randomly drawn from real persons with matching age and sex in the MTUS data set. We use a comparatively strict setting for data synthesis that focuses on the semantic accuracy of the created data. This means, we use a semantic similarity score weight of $\alpha = 0.9$ and a minimum similarity score threshold of $\tau_{sem} = 0.8$.

The data generation in ASTRA is based on an automated semantic mapping of activity descriptions from the input data set to POI descriptions in the simulation area. We replace this mapping approach with a fixed mapping that was created using a Large Language Model[1] and manual adaptations to be more realistic. After data generation, we filter out samples where the semantic mapping is erroneous. This can happen, if during data generation no suitable POIs were found due to spatiotemporal constraints. After pre-processing, the data set contains 55,650 locations from 7,822 users with 1 to 20 locations per user.

NCVR: The North Carolina Voter Registration (NCVR) data set (Christen, 2014) contains the home addresses and a set of sociodemographic attributes of registered voters in the US state of North Carolina. We first select a subset from NCVR based on a snapshot from August 2025, covering the city of Charlotte that represents an urban area. We then select a second subset from NCVR based on a snapshot from July 2024 that contains persons from zip code 27893 that is considered rural. We geocode all home addresses to obtain their GPS coordinates using Nominatim[2]. We use the demographic attributes *age, sex, race,* and *ethnicity*. The data set contains 310,202 locations in Charlotte and 20,520 locations in the rural zip code, which each correspond to one user.

4.2 Experimental Setup

We configure the attacks and evaluation as follows:

Privacy Protection: We employ two privacy mechanisms. First, as described in Section 2.2, we obfuscate the true locations of users with the planar Laplace mechanism (Andrés et al., 2013) to achieve ε-geo-indistinguishability using different privacy budgets ε. Second, we apply a generalization on the demographic and temporal attributes D and T of the users using three increasingly tight generalization levels. For more details of how we derive the relevant privacy parameters, ε and g, see Section 4.4.

Attack Parameters: To be able to compare the attack success for the different attacks and over regions with different population density, we use the same attack settings for all evaluations. We assume that for a wide range of applications an obfuscation distance of 5,000 m is the maximum value, that a user with strict privacy requirements would choose. We thus set $r_{max} = 5{,}000\,\text{m}$ accordingly. In many applications, an $\varepsilon < 1$ is considered to provide sufficient privacy, but often more relaxed values are used in practice[3]. Therefore, we set $\varepsilon_{max} = 5$.

We pre-evaluate which values of ε and g are required for a comprehensive evaluation in Section 4.4. Attacks that rely on an estimation of the attack circle's radius r_{attack} use the cumulative probability distribution of the Laplace function and the privacy budget ε to numerically calculate the radius at which 90% of obfuscations are located at most at this distance. The value is rounded to the next power of ten and a buffer of 20% is added. For example, a privacy budget of $\varepsilon = 0.001$ leads to $r_{attack} = 5{,}000\,\text{m}$, $\varepsilon = 0.01$ leads to $r_{attack} = 500\,\text{m}$, and so on.

Semantic Background Knowledge: For attacks that use a generic or a user-specific mobility profile (attacks ***PM***, ***PMD***, ***LE-PM***, and ***LE-PMD***), we calculate a probability distribution $Pr_{sem}(\cdot)$ based on the MTUS (Fisher et al., 2022; Gershuny et al., 2020) data set. This data set also functions as the source for ASTRA. It contains the daily activities of real persons along with the timestamps and their sociodemographic attributes.

We query the POIs P in R using OpenStreetMap[4] (OSM). We consider only POIs, where people typically spend their time and select POIs belonging to a subcategory of one of the OSM tags *amenity*, *building*, *office*, *shop*, *tourism*, *leisure*, and *sport*. We create a fixed mapping from the label of each POI $p \in P$ to the three most likely activities to be executed at this POI. For example, a POI *library* will be assigned to the activities *read*, *regular schooling and education*, and *homework*.

For each POI and its assigned activities their frequency of occurrence in the MTUS data set is summed and transformed into a likelihood, evenly spread over all POIs with the same label in the attack circle. If a user-specific mobility pattern is required, the data set is additionally filtered for persons matching the attacked person's sociodemographic features.

[1] OpenAI, ChatGPT (GPT-5.2), https://chat.openai.com

[2] https://nominatim.org

[3] https://desfontain.es/blog/real-world-differential-privacy.html

[4] https://www.openstreetmap.org

4.3 Evaluation Metrics

The success of an attack is commonly evaluated using the adversary's *expected estimation error* (Shokri et al., 2011; Wagner and Eckhoff, 2018). This metric evaluates the incorrectness of the adversary in identifying the true location x by calculating the Euclidean distance $d(\cdot)$ between the adversary's predicted location x^* and the true location x. It is calculated over the posterior probability of the adversary's predictions x^* given the observation x':

$$err_{exp}(x^*|x') = \sum_{x^*} Pr(x^*|x')d(x^* - x). \quad (6)$$

In the case that the adversary makes only one guess, this metric can be simplified to the *estimation error*:

$$err_d(x^*) = d(x^* - x) \quad (7)$$

The estimation error is a good indicator of the uncertainty of the adversary. However, the impact onto a user's privacy is particularly high when their true location can be pin-pointed, and even more if the semantic meaning can be identified correctly, indicating what the user was doing. We therefore additionally measure the *percentage incorrectly classified*, err_c (Wagner and Eckhoff, 2018), which indicates the ratio of incorrect predictions over all predictions:

$$err_c = \frac{1 - \sum_{x^*}(x \sim^c x^*)}{\sum_{x^*}} \quad (8)$$

where $\sim^c$ yields 1 if a prediction x^* is considered a correct match for x with regards to a certain condition c, or 0 otherwise. We define a prediction to be a correct match with regards to the location ($c = loc$), if the distance between the predicted location x^* and the true location x is less than 10 m or the true POI's geometry o contains x^*:

$$x \sim^{loc} x^* = (d(x,x^*) < 10\,\text{m}) \vee (x^* \subset o) \quad (9)$$

We define a prediction to be correct with regards to the POI ($c = poi$), if additionally the true POI's label l_p is correctly predicted:

$$x \sim^{poi} x^* = (x \sim^{loc} x^*) \wedge (l_p == l_{p^*}). \quad (10)$$

Here, $\sim^{poi}$ requires an exact match for all POI labels except labels *dormitory, apartments, residential, detached, semidetached house, house, terrace, yes* (referring to an unspecified building), or *apartment*, which are allowed to be used interchangeably as they all refer to a home POI where people live. In the case that the adversary cannot make a guess, for example because no suitable POI was found, the prediction is excluded from the metric calculations.

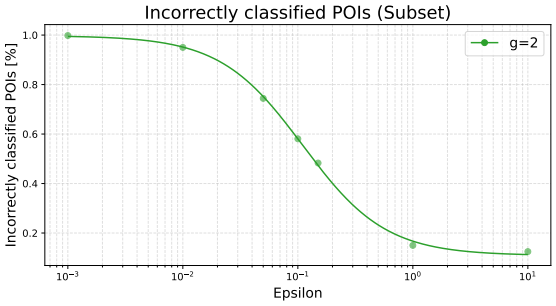


Figure 4: Percentage incorrectly classified POIs, err_{poi}, for varying privacy budgets ε and generalization level $g = 2$.

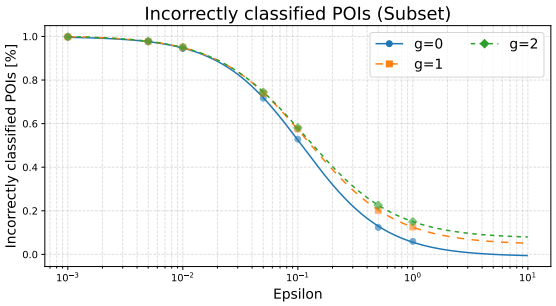


Figure 5: Percentage incorrectly classified POIs, err_{poi}, for the chosen privacy budgets ε and different generalization levels g.

4.4 Relevant Privacy Parameters

To find significant privacy parameter values for the obfuscation and generalization for the considered use case, we estimate them as follows. For both data sets, we sample a subset of locations that are located in a smaller subregion of 20x20 km containing both urban and rural areas. These subsets contain 16,268 locations for ASTRA and 44,092 locations for NCVR.

Privacy Budget ε: We define a search range of $\varepsilon \in [0.001, 10]$ and fix the generalization level at the coarsest resolution ($g = 2$, described in the next section). As described in Section 4.2, we estimate for each ε a suitable attack radius r_{attack}. We then run attack (LE-PMD) using all available demographic features and measure the attack success based on the average percentage of incorrectly classified POIs, err_{poi}. We fit a logistic curve to the data and calculate the inflection point. This point indicates the region of highest uncertainty and is chosen as the next ε. The process is repeated until the inflection point is stable within a small range.

Figure 4 shows the percentage of incorrectly classified POIs, err_{poi}, for the tested range of values of ε. Based on this evaluation we select the following values of ε for our experiments: $\varepsilon \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$

Table 2: Top features by lowest percentage of incorrectly classified POIs, err_{poi}.

Privacy Budget ε	Top features ASTRA	Top features NCVR
0.001	number of children (99.74%), single parenting (99.73%)	race (99.97%)
0.005	single parenting (97.63%), citizenship (97.6%)	ethnicity (99.68%)
0.01	income (94.7%), age (94.67%), season (94.63%)	ethnicity (98.88%)
0.05	student (74.7%), working time (74.18%), weekend (73.75%), number of children (73.53%), income (73.49%)	age (85.74%)
0.1	student (58.72%), weekend (58.24%)	ethnicity (64.38%)
0.5	weekend (22.82%)	ethnicity (2.77%)
1	age (13.16%)	age (0.4%)

leading to an estimated obfuscation distance of $r \in \{5{,}000\,\text{m}, 1{,}000\,\text{m}, 500\,\text{m}, 100\,\text{m}, 50\,\text{m}, 10\,\text{m}, 5\,\text{m}\}$.

Generalization Level g: To understand the impact of the generalization level on the attack success, we evaluate for each ε, derived in the previous step, the percentage of incorrectly classified POIs, err_{poi}, on the sample subset for three generalization levels $g \in \{0,1,2\}$. Higher generalization levels indicate increasing levels of privacy protection. For example, the value of the attribute *age* is transformed into one of 20 ($g = 0$), six ($g = 1$), or three ($g = 2$) non-overlapping intervals. The weekday (1 to 7, $g = 0$) is reduced to weekend status ($g \in \{1,2\}$), the season (spring to winter, $g = 0$) becomes indoor or outdoor season ($g \in \{1,2\}$), and so on.

Figure 5 shows err_{poi} for each generalization level g. As expected, a coarser generalization (higher generalization levels $g = 1$ and $g = 2$) increases the privacy protection (compared to $g = 0$) due to the reduction of information. To create a baseline for attacks using such demographic knowledge, we select the coarsest generalization level of $g = 2$ for all further evaluations.

4.5 Relevant Demographic Features

The goal of this pre-analysis is to identify the most significant demographic features to be used in attacks ***PMD*** and ***LE-PMD*** for each data set. We first carefully eliminate attributes that show a high correlation with other attributes. Second, we employ a forward feature selection approach, to identify the minimal set of attributes, yielding the lowest err_{poi}.

Based on the full data sets, we first calculate the correlation between each of the demographic attributes. Figure 6 shows the correlation matrix for ASTRA's demographic attributes. Due to a high correlation with the features *employed*, *retired* and *age*, we eliminate the feature *working time* from further analysis. We further eliminate *age of youngest child* because of its high correlation with *number of children*. For NCVR, there are no such high correlations between the attributes to be observed.

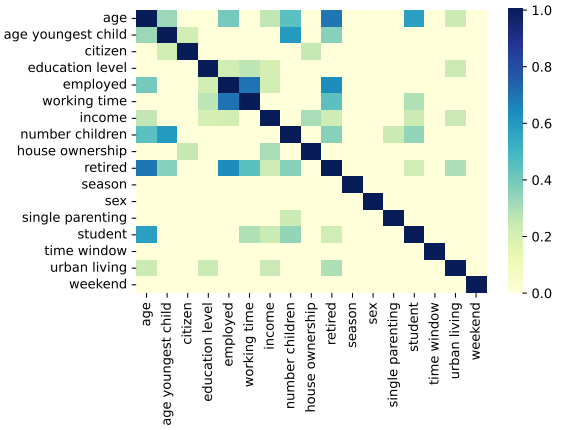
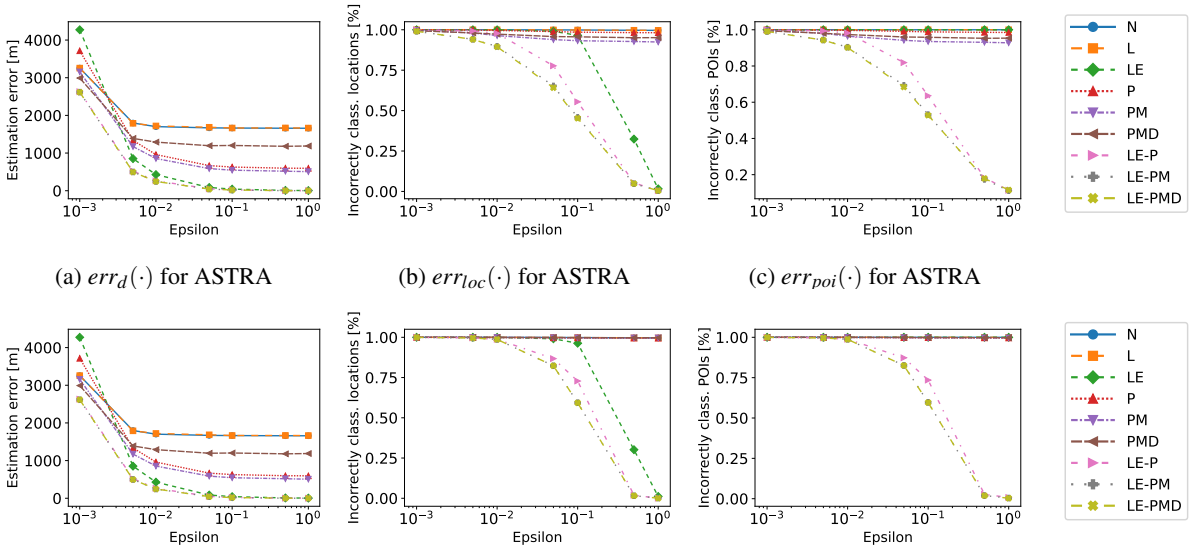


Figure 6: Absolute correlation of demographic features in ASTRA, thresholded by values of at least 0.2.

In the second step, we run the forward feature selection for each of the relevant privacy budgets ε, using generalization level $g = 2$, as we described in Section 4.4. First, we run attack ***LE-PMD*** with background knowledge comprising only one attribute. For the setting that yields the lowest err_{poi}, the attack is repeated with the respective attribute plus a second attribute. We continue this process until the attack success cannot be improved further by adding more attributes. Table 2 shows the top scoring features for the different privacy budgets ε for both data sets. For further analyses, we select those features that appear the most often in the top two for the different values of ε: {*weekend, student, age, single parenting*} for ASTRA and {*ethnicity*} for NCVR.

5 EXPERIMENTAL RESULTS

In this section, we compare the success of the nine different attack strategies to understand which characteristics of an attack and of the data pose the highest



(a) $err_d(\cdot)$ for ASTRA (b) $err_{loc}(\cdot)$ for ASTRA (c) $err_{poi}(\cdot)$ for ASTRA

(d) $err_d(\cdot)$ for NCVR (e) $err_{loc}(\cdot)$ for NCVR (f) $err_{poi}(\cdot)$ for NCVR

Figure 7: Error metrics across attack models for increasing privacy budgets ε.

privacy risk for individuals. Figure 7 shows the estimation error and the percentage of incorrectly classified locations and POIs for both data sets across the different attacks for increasing privacy budgets ε.

5.1 Impact of Privacy Knowledge

We first analyze the impact of the privacy knowledge on the attack success in both data sets. As expected, attack ***N***, which serves as a random baseline without any background knowledge, produces one of the highest estimation errors over all attacks. Additional knowledge of the privacy mechanism, as used in attack ***L***, can in most cases not improve the attack success. However, if the privacy budget ε is known, as in attack ***LE***, the adversary can significantly improve the estimation error for $\varepsilon > 0.001$. This is to be expected, because the knowledge of ε allows the adversary to make an accurate estimation of the obfuscation distance and they can adjust the attack radius accordingly. Similarly, the estimation error for attacks ***P***, ***PM***, and ***PMD***, relying solely on semantic background knowledge, can be significantly improved when the privacy budget ε is known in attacks ***LE-P***, ***LE-PM***, and ***LE-PMD***.

Knowing ε, however, cannot significantly reduce the percentage of incorrectly classified locations and POIs in attacks without semantic knowledge, indicating that an adversary is guessing in the correct area but still cannot identify the correct location. Overall, the privacy parameter has an important influence on the attack success and on the privacy of the user. While the results indicate that the disclosure of the used privacy mechanism does not increase the privacy risk for the user, the privacy budget should not be disclosed in cases where the Laplace mechanism is used to achieve geo-indistinguishability.

5.2 Impact of Semantic Knowledge

Next, we explore whether semantic background information is useful for an adversary to increase their attack success. Comparing the estimation error of attacks ***P***, ***PM***, and ***PMD***, which use only semantic knowledge, with the random baseline ***N***, a significant improvement can be observed for each attack. Attack ***PMD*** shows the smallest improvement of all three attacks for ASTRA, while for NCVR it is similar to the baseline. Also, the improvement is not as large as when the privacy budget ε is known as in attacks ***LE-P***, ***LE-PM***, and ***LE-PMD***. This indicates that semantic knowledge can be better leveraged when the obfuscation distance is well estimated and the adversary is searching in the correct area.

Attacks ***PM*** and ***LE-PM***, which consider typical mobility patterns of persons, can only slightly improve the estimation error compared to randomly picking any POI in the attack circle, as implemented by attacks ***P*** and ***LE-P***. The percentage of incorrectly classified locations and POIs, however, shows a larger improvement. This indicates that the prediction of POIs is more precise when the adversary is exploiting mobility patterns. Because semantic information of POI maps and mobility patterns of persons can be

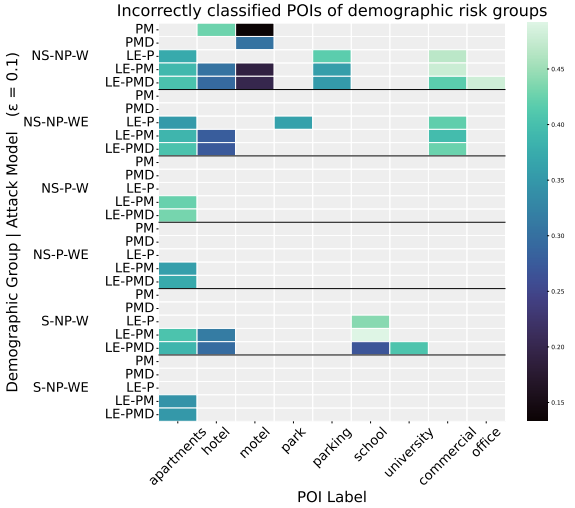


Figure 8: err_{poi} per demographic group, attack model, and POI label, where $err_{poi} < 50\%$ for at least 100 occurrences ($\varepsilon = 0.1$). Demographic groups are encoded as S/NS (Student/Non-student) - P/NP (Single parent/No single parent) - W/WE (Weekday/Weekend).

freely obtained by anyone, an assessment of the privacy risk involved in sharing location data in LBS-applications should therefore always take into account such information.

Opposed to our expectation, attacks ***PMD*** and ***LE-PMD***, incorporating user-specific mobility patterns cannot further improve the results for any of the three metrics compared to using a generic mobility profile as done by attacks ***PM*** and ***LE-PM***. For attack ***PMD*** the success is even worse. However, the success can be elevated for certain demographic groups using this attack, as we describe in Section 5.4.

5.3 Impact of Combined Knowledge

When combining both privacy and semantic background knowledge in attacks ***LE-P***, ***LE-PM***, and ***LE-PMD***, it is not surprising that the highest attack success can be achieved for all three metrics. In particular, the percentage of incorrectly classified locations and POIs are significantly reduced down to error rates close to 0% for the three attacks. The effect is more prominent for privacy budgets of $\varepsilon > 0.01$. This indicates, that synergy effects appear when both semantic and privacy knowledge are available for an attack.

5.4 Impact of Demographic Features and POI Label

While attacks ***PMD*** and ***LE-PMD*** cannot generally improve the error metrics, there can be a variation

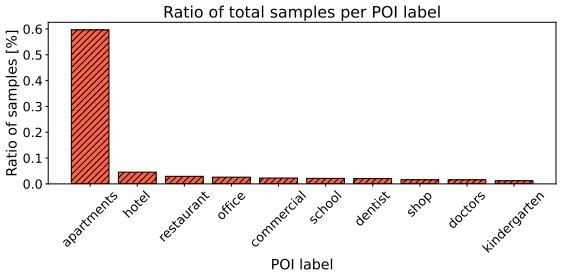


Figure 9: Ratio of top ten POI labels in ASTRA.

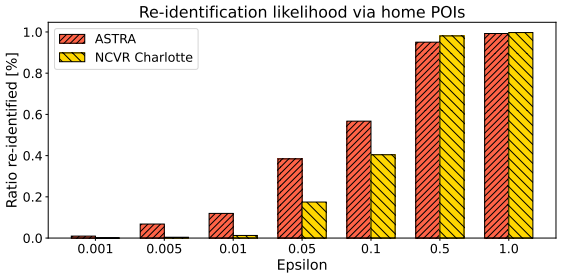
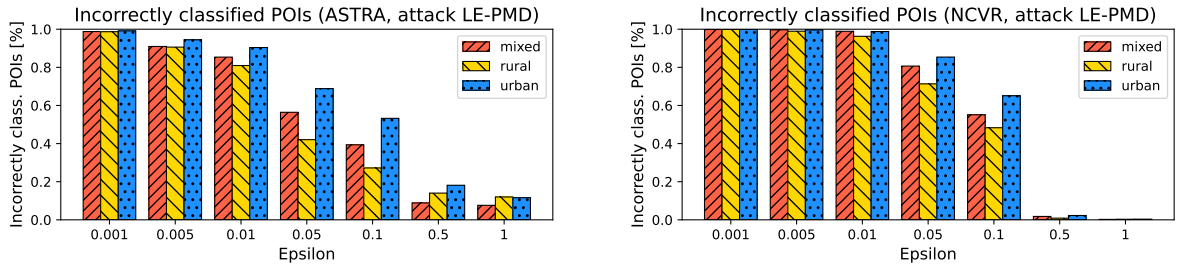


Figure 10: Re-identification likelihood exposed by home locations using attack ***LE-PMD***.

between individual demographic groups or for certain POI labels. We thus analyze the privacy risk for such groups as follows. Based on the significant demographic features for ASTRA (see Section 4.5), *student*, *single parenting*, and *weekend*, we create groups from each combination (for better readability, we leave out *age*). We aggregate the results per demographic group, attack model, and POI label, and filter out combinations that have less than 100 occurrences and an average $err_{poi} < 50\%$. Figure 8 shows the results for $\varepsilon = 0.1$ in the ASTRA data set.

Only the five attacks with the most background knowledge (***PM***, ***PMD***, ***LE-P***, ***LE-PM***, and ***LE-PMD***) display significant results. The figure indicates that different home locations (which we summarized as *apartments*), can be identified well among all demographic groups. However, this is also the POI label that occurs with 60% the most often in the ASTRA data set (see Figure 9). This POI type is especially sensitive as it might re-identify a user when their home location becomes known. The re-identification risk from such POIs through attack ***LE-PMD*** can be substantial, if ε is not chosen carefully. Figure 10 shows the ratio of re-identified home locations, which is calculated as the percentage of correctly classified POIs, $(1 - err_{poi})$ (based only on home-related locations), and weighted by the ratio of successful predictions over these POIs.

Hotels and motels are also well identified, even though their ratio of samples is small. This could be caused by such POIs being rare and thus easier to



(a) ASTRA (b) NCVR

Figure 11: err_{poi} per population density for ASTRA and NCVR for attacks over all privacy budgets ε.

identify. The same is true for offices and commercial places where persons work. This risk is mainly incurring for users from the non-student group, which is largely comprised of workers. For students (which includes pupils) there is an increased risk on weekdays that educational institutions, such as universities or schools, are identified by an attack.

Note that these results are underlying the given distribution of demographic attributes and POI labels in the ASTRA data set and our mapping approach, as we described in Section 4.2.

5.5 Impact of Population Density

To analyze the impact of the population density on the attack success, we divide the geographical space R into square cells of 1,000 m cell width and assign each grid cell a label, based on the number of inhabitants per square kilometer as follows: *urban* (more than 1,500), *rural* (less than 300), or *mixed* (else). Figure 11 shows err_{poi} per population density for ASTRA and NCVR for attack ***LE-PMD*** over all privacy budgets ε. The lowest error rates can be observed for rural areas, while locations in urban areas are the hardest to reconstruct. This is to be expected, because the density of POIs is higher in urban areas, increasing the number of potential POIs to choose as prediction. The impact is more significant for increasing values of ε up to $\varepsilon = 0.1$.

6 DISCUSSION

We now discuss the key findings of our work, its limitations and future work.

6.1 Key Findings

In our experimental evaluation we investigated potential privacy leaks through context linking attacks in an LBS scenario. We further studied the impact that different levels of background knowledge can have onto the success of localization attacks. Our results indicate that an adversary with sufficient background knowledge can in fact reconstruct a user's true location from an obfuscated one, and even find out their semantic meaning. The highest attack success can be achieved when knowledge of the privacy mechanism (and in particular the privacy parameter ε) is combined with semantic knowledge. Semantic information, such as POIs, have a high influence on the results because they significantly limit the possible locations that a user can be at. Without a rough estimate, however, of the obfuscation distance, the attack success is not as significant.

Against our expectation, an adversary who exploits a user's demographic features cannot increase their attack success in every case. However, for distinct demographic groups, such as students, the privacy risk can be higher as their whereabouts are more predictable. Confirming our expectation, locations in rural areas are more likely to be reconstructed than in urban areas.

These results need to be taken into account when designing new location privacy-preserving mechanisms. While related works integrate such background knowledge in their protection schemes, they require a structured approach considering such knowledge depending on its influence on privacy leaks. Our work is an initial step towards such a structured analysis of background knowledge. Users of LBS or similar services need to be aware of whether they belong to a risk group, which can increase their risk of privacy leakage.

6.2 Limitations and Future Work

Our analysis provides first indications as to which types of background knowledge carry the highest privacy risk, but has certain limitations. While our ASTRA data set (Schneider et al., 2026) is real-world based, it might be limited with regards to the demo-

graphic and POI variety and is underlying assumptions in the data generation process. Our results on the ASTRA data set are backed by the real-world NCVR data set (Christen, 2014), which is however limited to only home-related POIs. The generalizability of our results can thus be affected to over- or underestimate the risk for certain demographic groups and POIs to be identified. Future work should re-evaluate our findings using large real-world data sets with greater geographical and semantic diversity.

While our risk analysis focuses specifically on the Laplace mechanism, other DP-based protection mechanisms remain to be investigated. Future work should further investigate the privacy risk under more elaborate attack approaches. For example, in an improved attack, an adversary can consider typical POI visit patterns and compare them to the actual POI distribution around the observed user location. Because urban morphology is very unique, this approach can help to pinpoint more likely areas for a user to be in (Cao et al., 2018). Advanced attacks should further leverage the capabilities of deep learning techniques to better model the mobility patterns in large training data sets (Buchholz et al., 2022).

While in our work we assume that a user is sending a single request to the LBS, further studies should investigate the privacy risk from repeated requests. This might allow the adversary to estimate a more specific mobility pattern for a single user. Because such data points are often correlated (for example, when sharing trajectory data), the adversary might be able to acquire a more accurate probability distribution, indicating where a user is present at a given time.

6.3 Ethical Considerations

For our experiments, we stored and processed all datasets exclusively on local devices and secured networks to meet ethical requirements. We explicitly did not attempt to re-identify any individuals or vulnerable groups in the datasets. Furthermore, we disclose only aggregate statistical performance measures and do not release any individual-level or raw data. Consequently, all experiments were conducted in compliance with institutional and legal requirements.

7 CONCLUSION

In this work, we examined whether differentially private locations of individuals can be reconstructed via context linking attacks and which type of background knowledge causes the highest privacy leak. Our results show that a successful reconstruction is in fact possible with almost 50% of correctly guessed POIs for a privacy budget of $\varepsilon = 0.1$, which relates to an obfuscation of about 50 m. In particular, semantic background information about POIs and mobility patterns of humans significantly increase the privacy risk caused by such attacks, especially when combined with knowledge about the chosen parameters of the privacy mechanism.

Our work highlights which pieces of adversarial knowledge should be taken into account the most when devising semantic-aware location protection mechanisms and thus contributes to an effective protection of human mobility data. While our work provides initial findings, future work should investigate in more depths different types of background knowledge, more fine-granular demographic features and more advanced attack approaches, using real-world data sets with larger demographic variety.

ACKNOWLEDGEMENTS

This work was partially funded by Universities Australia and the German Academic Exchange Service (DAAD) grant 57701258.

REFERENCES

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., and Palamidessi, C. (2013). Geo-indistinguishability: Differential privacy for location-based systems. *Proceedings of the ACM Conference on Computer and Communications Security*, pages 901–914.

Buchholz, E., Abuadbba, A., Wang, S., Nepal, S., and Kanhere, S. S. (2022). Reconstruction Attack on Differential Private Trajectory Protection Mechanisms. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 279–292.

Cao, H., Feng, J., Li, Y., and Kostakos, V. (2018). Uniqueness in the City: Urban Morphology and Location Privacy. *Ubicomp*, 2(2):1–20.

Chatzikokolakis, K., Palamidessi, C., and Stronati, M. (2015). Constructing elastic distinguishability metrics for location privacy. *Proceedings on Privacy Enhancing Technologies*, 2015(2):156–170.

Christen, P. (2014). Preparation of a real temporal voter data set for record linkage and duplicate detection research. Technical report, Australian National University.

Dwork, C. (2011). A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science*, pages 265–284. Springer, Berlin, Heidelberg.

Fisher, K., Gershuny, J., Flood, S. M., Backman, D., Vega-Rapun, M., Lamote, J., and Sayer, L. C. (2022). Multinational Time Use Study Extract System: Version 1.4 [dataset]. Minneapolis, MN: IPUMS.

Gershuny, J., Vega-Rapun, M., and Lamote, J. (2020). Multinational Time Use Study [dataset] Centre for Time Use Research, UCL IOE, University College London [www.timeuse.org/mtus/].

Guerra-Balboa, P., Pascual, A. M., Parra-Arnau, J., Forné, J., and Strufe, T. (2022). Anonymizing Trajectory Data: Limitations and Opportunities. In *Proceedings of the Third AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI-22), Virtual*, volume 28.

Hoh, B., Gruteser, M., Hui, X., and Alrabady, A. (2006). Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., and Varshavsky, A. (2011). Identifying Important Places in People's Lives from Cellular Network Data. In *Pervasive Computing*, volume 6696, pages 133–151. Springer Berlin Heidelberg, Berlin, Heidelberg.

Krumm, J. (2007). Inference attacks on location tracks. In *International Conference on Pervasive Computing*, pages 127–143. Springer Berlin Heidelberg.

Li, H., Zhu, H., Du, S., Liang, X., and Shen, X. (2018). Privacy Leakage of Location Sharing in Mobile Social Networks: Attacks and Defense. *IEEE Transactions on Dependable and Secure Computing*, 15(4):646–660.

Liao, L., Fox, D., and Kautz, H. (2005). Location-based activity recognition using Relational Markov Networks. *IJCAI International Joint Conference on Artificial Intelligence*, pages 773–778.

Liu, B., Zhou, W., Zhu, T., Gao, L., and Xiang, Y. (2018). Location Privacy and Its Applications: A Systematic Study. *IEEE Access*, 6:17606–17624.

Mahmud, J., Nichols, J., and Drews, C. (2014). Home location identification of twitter users. *ACM Transactions on Intelligent Systems and Technology*, 5(3).

Maouche, M., Mokhtar, S. B., and Bouchenak, S. (2017). AP-Attack: A Novel User Re-identification Attack On Mobility Datasets. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 48–57, Melbourne VIC Australia. ACM.

Peng, Z., An, J., Gui, X., Wang, Z., Zhang, W., Gui, R., and Xu, J. (2019). Location Correlated Differential Privacy Protection Based on Mobile Feature Analysis. *IEEE Access*, 7:54483–54496.

Primault, V., Mokhtar, S. B., Lauradoux, C., and Brunie, L. (2014). Differentially Private Location Privacy in Practice. In *Proceedings of the Third Workshop on Mobile Security Technologies (MoST) 2014*, pages 1–10.

Sadilek, A., Kautz, H., and Bigham, J. P. (2012). Finding your friends and following them to where you are. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 723–732, Seattle Washington USA. ACM.

Samarati, P. and Sweeney, L. (1998). Generalizing Data to Provide Anonymity when Disclosing Information. In *ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, pages 10–1145.

Schneider, M., Buchmann, E., and Rahm, E. (2024). Distributed, Privacy-Aware Location Data Aggregation. In *2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*, pages 31–40, Washington, DC, USA. IEEE.

Schneider, M., Nanayakkara, C., Mohn, M., Christen, P., and Rahm, E. (2026). Generating Semantically Enriched Mobility Data from Travel Diaries. In *Advances in Databases and Information Systems: 29th European Conference, ADBIS 2025, Tampere, Finland, September 23–26, 2025, Proceedings*, volume 16043 of *Lecture Notes in Computer Science*, Cham. Springer Nature Switzerland.

Shen, Z., Lu, S., Huang, H., Yuan, M., Tang, G., Chen, W., Zhang, T., and Zhong, T. (2020). An Approach Based on Customized Robust Cloaked Region for Geographic Location Information Privacy Protection. *Mobile Information Systems*, 2020:1–12.

Shokri, R., Theodorakopoulos, G., Le Boudec, J. Y., and Hubaux, J. P. (2011). Quantifying location privacy. *Proceedings - IEEE Symposium on Security and Privacy*, pages 247–262.

Tian, C., Xu, H., Lu, T., Jiang, R., and Kuang, Y. (2021). Semantic and Trade-Off Aware Location Privacy Protection in Road Networks Via Improved Multi-Objective Particle Swarm Optimization. *IEEE Access*, 9:54264–54275.

Wagner, I. and Eckhoff, D. (2018). Technical privacy metrics: A systematic survey. *ACM Computing Surveys*, 51(3).

Wernke, M., Skvortsov, P., Dürr, F., and Rothermel, K. (2014). A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing*, 18(1):163–175.

Wu, L., Yang, L., Huang, Z., Wang, Y., Chai, Y., Peng, X., and Liu, Y. (2019). Inferring demographics from human trajectories and geographical context. *Computers, Environment and Urban Systems*, 77:101368.

Yan, Y., Xu, F., Mahmood, A., Dong, Z., and Sheng, Q. Z. (2022). Perturb and optimize users' location privacy using geo-indistinguishability and location semantics. *Scientific Reports*, 12(1):20445.

Zhang, S., Li, M., Liang, W., Sandor, V. K. A., and Li, X. (2022). A Survey of Dummy-Based Location Privacy Protection Techniques for Location-Based Services. *Sensors*, 22(16):6141.

Zhong, Y., Yuan, N. J., Zhong, W., Zhang, F., and Xie, X. (2015). You are where you go: Inferring demographic attributes from location check-ins. *WSDM'15*, pages 295–304.