Membership Inference Attacks Against Indoor Location Models

Vahideh Moghtadaiee[1], Amir Fathalizadeh[1] and Mina Alishahi[2]
[1]*Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran*
[2]*Department of Computer Science, Open Universiteit, Amsterdam, The Netherlands*

Keywords: Membership Inference Attack, Indoor Localization, Differential Privacy, Location Privacy.

Abstract: With the widespread adoption of location-based services and the increasing demand for indoor positioning systems, the need to protect indoor location privacy has become crucial. One metric used to assess a dataset's resistance against leaking individuals' information is the Membership Inference Attack (MIA). In this paper, we provide a comprehensive examination of MIA on indoor location privacy, evaluating their effectiveness in extracting sensitive information about individuals' locations. We investigate the vulnerability of indoor location datasets under white-box and black-box attack settings. Additionally, we analyze MIA results after employing Differential Privacy (DP) to privatize the original indoor location training data. Our findings demonstrate that DP can act as a defense mechanism, especially against black-box MIA, reducing the efficiency of MIA on indoor location models. We conduct extensive experimental tests on three real-world indoor localization datasets to assess MIA in terms of the model architecture, the nature of the data, and the specific characteristics of the training datasets.

1 INTRODUCTION

Privacy in indoor localization is a major concern as it involves tracking people within indoor spaces, potentially revealing their social habits, interests, workplaces, and health status (Alhalafi and Veeraraghavan, 2019; Fathalizadeh et al., 2023). Unlike GPS-based outdoor localization, indoor positioning systems using Wi-Fi, Bluetooth, etc. lack robust privacy safeguards, raising concerns about unauthorized access, tracking without consent, targeted advertising, and malicious activities (Sartayeva and Chan, 2023; Fathalizadeh et al., 2024). Accumulating location data reveals behavior patterns, emphasizing the need for privacy measures in ethical deployment (Navidan et al., 2022).

One prominent privacy attack is the membership inference attack (MIA), where adversaries try to determine if a specific record was part of the training data for a machine learning (ML) model trained on sensitive information, such as indoor location data. MIA exploits information leakage to identify whether a data point was included in the training set, raising privacy concerns for sensitive applications (Shokri et al., 2017). In cases involving sensitive data like medical records, financial data, or location information, the consequences of an MIA breach can be severe (Dionysiou and Athanasopoulos, 2023). Privacy-preserving techniques like differential privacy (DP), anonymization, and generative models aim to protect data privacy from attackers. However, studies show that these techniques can still be vulnerable to MIA, posing risks to the confidentiality of the training data (Hyeong et al., 2022). Among these methods, DP is widely used, offering probabilistic protection during ML model training. Models trained with DP do not retain specific records, potentially shielding against MIA. This underscores the importance of DP in mitigating privacy risks.

In localization context, MIA can reveal if specific locations or individuals were part of a model's training data, risking unauthorized tracking and privacy breaches (Zhang et al., 2020). Indoor datasets, unlike outdoor ones, have higher spatial granularity and distinct patterns due to confined spaces and a higher density of Access Points (APs) (number of features), necessitating careful MIA assessment. These differences drive our investigation into MIA vulnerabilities in indoor localization models. This paper presents the first evaluation of MIA on indoor localization systems, including assessments after privatizing the training dataset with DP. We also examine two possible MIA attack models before and after privatizing indoor location data: *i*) *white-box*, where the adversary knows the target model and parameters, and *ii*) *black-box*, where the adversary lacks such knowledge. The study evalu-

Moghtadaiee, V., Fathalizadeh, A. and Alishahi, M.
Membership Inference Attacks Against Indoor Location Models.
DOI: 10.5220/0012863100003767

In *Proceedings of the 21st International Conference on Security and Cryptography (SECRYPT 2024)*, pages 584-591
ISBN: 978-989-758-709-2; ISSN: 2184-7711

ates MIA performance and potential privacy leakage using three real-world indoor location datasets.

The paper is structured as follows: Section 2 reviews related work, Section 3 covers preliminary concepts, and Section 4 introduces the MIA architecture in indoor localization and adversary methodology. The experimental setup and evaluation are in Section 5, and Section 6 concludes the paper.

2 RELATED WORK

The first MIA against classification models under ML-as-a-service (MLaaS) is introduced by (Shokri et al., 2017). They measure leaked membership information and propose a shadow training technique to perform membership inference based solely on the model's outputs. However, their approach is limited by the assumption that the target and shadow models have consistent structures and training data. Subsequent studies (Salem et al., 2018) and (Carlini et al., 2022), extend MIA techniques. Authors in (Salem et al., 2018) relax adversarial assumptions and evaluate membership privacy, proposing defenses like dropout and model stacking. A recent survey provides an overview of MIA literature on both attacks and defenses (Hu et al., 2022). DP counters MIA by adding noise to the objective function or gradients during training. Authors in (Hyeong et al., 2022) assess tabular data synthesis models' vulnerability to MIA and the effectiveness of DP-SGD and DP-GAN in mitigating it. However, current defenses often affect outputs rather than preserving the privacy of source data (Yang et al., 2023).

Regarding MIA on location data models, authors in (Shokri et al., 2017; Salem et al., 2018; Rahimian et al., 2020; Hui et al., 2021; Choquette-Choo et al., 2021; Liu et al., 2022) employ MIA methodologies to evaluate location data models Specifically, (Zhang et al., 2020) advances this problem by exploring more realistic attack scenarios. However, these studies focus on outdoor location ML models. While progress has been made in membership inference for location-based data, a notable gap exists in addressing MIA for indoor location data. Previous studies have not explored MIA against ML models trained on indoor location datasets, despite the unique challenges posed by indoor datasets, including higher spatial granularity, distinct signal patterns, and increased AP density. Additionally, while (Fathalizadeh et al., 2023) discusses the impact of DP on indoor localization accuracy, the application of DP to the original location dataset before model training and subsequent evaluation of MIA efficiency on this model remain unexplored.

3 PRELIMINARIES

3.1 Indoor Localization

Location fingerprinting draws inspiration from the unique fingerprints of individuals, with each location exhibiting distinct signal characteristics, often manifested as Received Signal Strength (RSS) values from Wi-Fi signals, commonly used in indoor localization (Hayward et al., 2022). The process involves two main stages: tra9ning and localization, both relying on Wi-Fi APs as signal references. During training, data collected through surveying the indoor area is used to create a radiomap containing RSS measurements or other features from known Reference Points (RPs), with their associated (x,y) coordinates or zone number z stored in a dataset called a radiomap on the server. Additionally, data collection can also be conducted through extensive data collection from individuals in the targeted area via crowdsourcing (Alikhani et al., 2018). In the localization stage, systems analyze signals from APs to predict user location, employing techniques such as traditional ML algorithms, Deep Learning (DL), and deep reinforcement learning (Roy and Chowdhury, 2021).

3.2 Differential Privacy

Differential privacy (DP) is designed for scenarios with a trusted data curator who collects data from individuals, processes it to satisfy DP constraints, and releases results (Dwork et al., 2006). DP limits the impact of a single data point on the overall output.

Definition 1 ((ε,δ)-Differential Privacy (Dwork et al., 2006))**.** *An algorithm $\mathcal{M}$ satisfies (ε,δ)-differential privacy ((ε,δ)-DP), where $\varepsilon > 0$, $\delta \geq 0$, if and only if for any two neighboring datasets D and D':*

$$\forall T \subseteq Range(\mathcal{M}) : Pr[\mathcal{M}(D) \in T] \leq e^{\varepsilon} Pr[\mathcal{M}(D') \in T] + \delta,$$

where $Range(\mathcal{M})$ denotes the set of all possible outputs of the algorithm $\mathcal{M}$. Two datasets D and D' are considered neighbors ($D \sim D'$), if either $D = D' + r$ or $D' = D + r$, where $D + r$ represents the dataset obtained by adding the record r to dataset D.

PrivSyn (Zhang et al., 2021) is a DP algorithm for synthesizing tabular microdata for data analysis. It generates a synthetic dataset D_2 from an original dataset D_1, ensuring statistical similarity between them. Formally, D consists of n records with k attributes. D_2 is considered similar to D_1 if $f(D_2)$ closely approximates $f(D_1)$ for any function f. *PrivSyn* focuses on three statistical measures: marginal queries, range queries, and classification models. Marginal queries

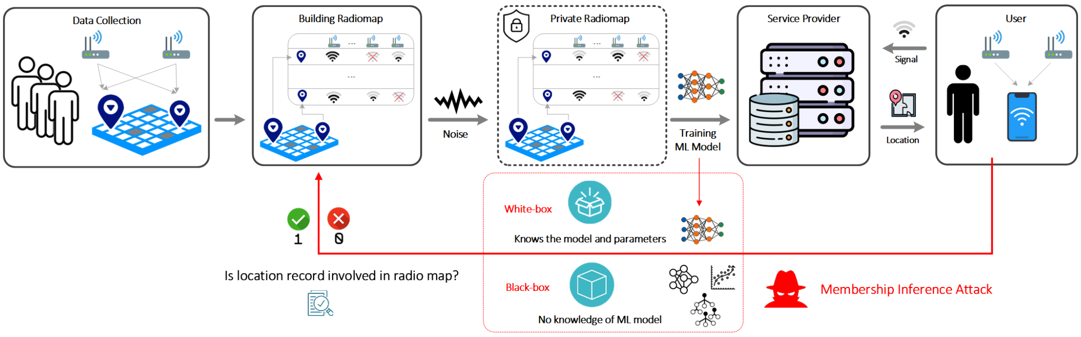


Figure 1: MIA architecture against differentially private indoor location models.

capture joint attribute distributions, range queries count records falling within specified ranges, and classification models are trained and evaluated using the synthetic dataset.

3.3 Membership Inference Attack (MIA)

MIA aims to detect if a data point is part of a target model's training dataset, risking privacy by potentially exposing raw data and model details (Hyeong et al., 2022). It uses methods like analyzing model confidence scores, probing decision boundaries, and exploiting vulnerabilities. Attackers balance efficiency and detectability when choosing techniques. MIA's significance lies in its potential to breach privacy as ML models increasingly handle sensitive data, with its easy deployment amplifying its importance. Given a target data point x_T, a target ML model $\mathcal{M}$ trained on the original dataset $\mathcal{D}$, and adversary prior knowledge denoted as I, MIA relies on training a binary classifier to understand an individual's membership status as a member or non-member.

4 MIA AGAINST INDOOR LOCATION MODEL

4.1 Problem Formulation

This paper focuses on indoor localization as an ML classification problem, aiming to determine the specific zone within a building where a user is located. Each zone (z) represents a distinct label in the classification task, typically corresponding to different rooms, corridors, or hallways. For instance, zones could be labeled as 'Room A', 'Room B', 'Corridor 1', 'Corridor 2', etc. The task involves training a classification model to predict the zone accurately based on RSS values or other features collected from APs deployed indoors. In indoor positioning, a classifier, denoted as model $\mathcal{M}$, takes an input X, *e.g.*, a vector of RSS values (as features), and estimates an output vector as $\mathcal{M}(\mathrm{X}) = \mathrm{Y}$. The length of Y corresponds to the number of class labels (zones), and each element represents the probability of each zone. The attack model $\mathcal{A}$ is constructed based on the top posterior probabilities, $Pr(Y|X)$, sorted in descending order. If the highest probability surpasses a predefined threshold, the location is categorized as a member of the training process; otherwise, it is classified as a non-member (Rahimian et al., 2020) as below:

$$\mathcal{A} = \begin{cases} 1 & \max Pr(Y|X) \geq \text{threshold} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

4.2 Attack Methodology

The architecture of MIA targeting the model trained on indoor location datasets is illustrated in Fig.1. As observed, the location data collected from individuals is used to construct a radiomap for indoor location inference. To enhance privacy, noise is injected into this radiomap, ensuring DP while maintaining accuracy. Next, the differentially private radiomap serves as the training dataset for the indoor location model, stored by the service provider. Users query the provider for their locations, with the trained model using the radiomap to predict based on their RSS values. However, attackers may also attempt MIA by querying the service provider. Attack model $\mathcal{A}$ aims to determine if a user or location was part of the training process. In white-box settings, attackers with model knowledge use sophisticated strategies, while black-box settings require alternative inference methods due to limited information.

To implement the attack, the adversary trains a shadow model to mimic the target model's behavior, using it to gather ground truth location data for training model $\mathcal{A}$. This requires a dataset, $\mathcal{D}_s$, from the same distribution as the target model's training data, which can be obtained by querying the server with sample RSS inputs. The shadow model construction varies by attack type: in a white-box setting, the adversary has access to the model's parameters and internal structure, while in a black-box setting, only the model's outputs are accessible. Both settings are common in practice (Song et al., 2023) and are detailed in the following subsections.

4.3 White-Box Setting

In the white-box attack model, the adversary is aware of the target model and its associated hyperparameters, relying on this transparency for informed inference. The attacker's objective in the white-box setting is expressed as follows:

$$\text{Given } \mathcal{M}(\cdot), \mathcal{D}, I \text{ determine if } \mathcal{M}(\mathrm{x}_T) \text{ reveals } \mathrm{x}_T \in \mathcal{D} \tag{2}$$

To conduct the white-box attack, we follow the procedure described in (Salem et al., 2018). The adversary is assumed to have a shadow dataset $\mathcal{D}_s$, which is divided into a shadow training dataset $\mathcal{D}_s^t$ and an output shadow dataset $\mathcal{D}_s^o$. The attacker uses $\mathcal{D}_s^t$ to train the shadow model $\mathcal{S}$ and then employs $\mathcal{S}$ to make predictions on $\mathcal{D}_s$ (comprising $\mathcal{D}_s^t$ and $\mathcal{D}_s^o$), obtaining posterior probabilities for each data point. For each data point, the highest posterior probabilities form its feature vector. A feature vector is labeled as 1 (member) if its corresponding data point is in $\mathcal{D}_s^t$, and 0 (non-member) otherwise. These feature vectors and labels are then used to train the attack model $\mathcal{A}$. To determine if a target x_T is in the training dataset, the adversary queries $\mathcal{M}$ with x_T to obtain its posterior probabilities, selects the maximum probabilities (sorted highest to lowest) and inputs them into $\mathcal{A}$ for membership prediction.

4.4 Black-Box Setting

The black-box attack model is more challenging, as attackers lack explicit knowledge of the target model. The attacker's objective in the black-box setting is:

$$\text{Given } \mathcal{M}(\mathrm{x}_T), \text{determine if } \mathrm{x}_T \in \mathcal{D}. \tag{3}$$

To implement the black-box attack, we use a collection of classifiers as shadow models to attack the target model, enhancing the attack by combining multiple ML models, each using a different classification algorithm, as described in (Salem et al., 2018). These models form a comprehensive shadow model, with each sub-shadow model trained on the same data. The features from all sub-shadow models are combined to create a larger dataset for training the attack model $\mathcal{A}$. This approach helps the shadow model understand various classifiers' behaviors, facilitating an attack on an unknown target model, assuming one sub-shadow model matches the target model's classifier.

5 EXPERIMENTS AND RESULTS

5.1 Datasets Description

CRI: The experimental dataset is a $51m \times 18m$ testbed on the second floor of the Cyberspace Research Institute at Shahid Beheshti University, containing 9 APs and 384 RPs. RSS values from all APs are measured at RPs in four directions, with 100 samples per direction, and stored in the radiomap. Further details are available in (Moghtadaiee et al., 2019).

JUIndoorLoc: The database covers a building with five floors situated at Jadavpur University (Roy et al., 2019). Each floor spans an area of $882m^2$, and the entire space is subdivided into grids measuring $1m^2$. Our focus is the fourth floor, and we consider 24 APs and 645 RPs within this floor.

UJIndoorLoc: This is an indoor location dataset including multiple buildings and stories of Universitat Jaume I, utilizing Wi-Fi fingerprints. Details regarding this dataset can be found in (Torres-Sospedra et al., 2014). We select a diverse area on the first floor of building No. 1, comprising a total of 198 RPs out of 5249 RPs across the entire building complex. Within the chosen area, 41 APs are detected in various locations, with 18 APs consistently sensed at all RPs.

5.2 Evaluation Metrics

We evaluate membership inference using precision, recall, and attack accuracy, adhering to the metrics outlined in (Zhang et al., 2020). In these metrics, True Positive (TP) represents correctly predicted 'member' instances, while True Negative (TN) denotes correctly predicted 'non-member' instances. False Positive (FP) indicates 'non-member' instances incorrectly classified as 'member', and False Negative (FN) represents 'member' instances incorrectly classified as 'non-member'.

Precision: Precision is the fraction of the specific data points predicted as members of the datasets used for training the ML model. It is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall: Recall represents the fraction of the specific data points that can be correctly inferred as members, which measures the coverage of our attack. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Attack Accuracy: It can be used as the privacy metric to evaluate the privacy-preserving performance. The

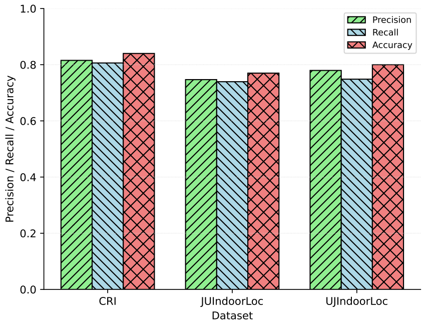


Figure 2: Comparision of White-box MIA's performance for each dataset.

higher the attack accuracy, the more serious the privacy leak and the more successful the attack is. The attack accuracy is calculated as follows:

$$AttackAccuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

5.3 Experimental Results

We evaluate MIA's privacy impact on indoor location datasets from various angles. Effectiveness varies with model architecture, data traits, and training dataset attributes, including pre-training DP application. Our simulation follows the original code configuration in (Salem et al., 2018)[1] and (Shokri et al., 2017)[2].

5.3.1 Overall MIA Performance in White-Box

In the white-box scenario, where the adversary knows the model used to train the dataset, we employ a multilayer perceptron (MLP) as the foundational training model, assuming the adversary is aware of this choice. The MLP used as the target model has a single hidden layer with 128 units. Each shadow model replicates the architecture of its corresponding target model. Additionally, we construct the attack model using another MLP model with a hidden layer containing 64 units and a softmax output layer. The outcomes of this experiment are illustrated in Fig. 2. Results indicate that MIA has a 74% success across all datasets.

5.3.2 Impact of Epochs and Shadow Models Number in White-Box Setting

To examine the effects of epoch number and shadow model size, we varied number of epochs from 10 to 150 and shadow models from 1 to 20. Fig. 3 and 4 present precision, recall, and attack accuracy metrics across three datasets. To isolate individual variables' effects, we fix shadow models at 10 in Fig. 3 and epochs at 100 in Fig. 4. Both variables show similar trends, with increasing epochs and shadow models enhancing MIA success up to a threshold (identified as 80 epochs and 10 shadow models), beyond which further increments yield marginal improvements.

5.3.3 The Impact of Dataset Properties on MIA Performance in White-Box Setting

Our objective here is to assess how adjustments to input dataset parameters, such as dataset size, number of zones (class labels), and number of APs (features), affect MIA accuracy.

Dataset Size: Fig. 5a depicts dataset size impact on MIA accuracy as the percentile fraction changes. The numbers of APs and zones are kept at their maximum values in the original dataset. Remarkably, larger datasets diminish MIA effectiveness across all datasets. This is due to the increased difficulty in identifying specific records from a larger pool.

Number of APs: Fig. 5b shows how varying the number of APs (features) impacts attack accuracy across datasets. An increased number of features leads to reduced accuracy. This is because fewer features create a more concentrated set of characteristics that the model relies on, potentially aiding attackers in identifying specific data points indicating higher vulnerability to MIA.

Number of Zones: Fig. 5c illustrates how the number of zones affects MIA accuracy for the CRI dataset. Increasing zone count leads to MIA accuracy decline, as the model captures more nuanced indoor environment patterns. Higher zone numbers introduce complexity and diversity, making it harder for attackers to infer membership status accurately.

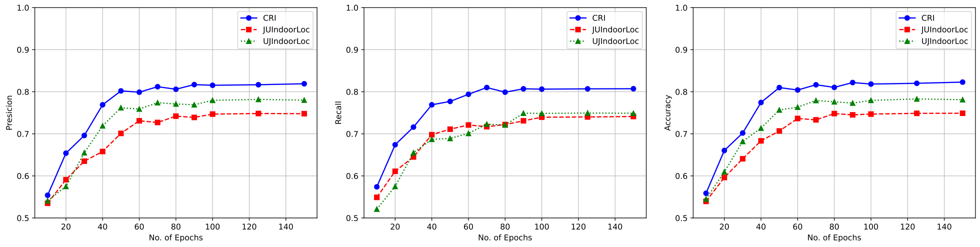
These findings underscore the critical impact of variables such as the number of users, features, and zones on the success of MIA attacks. While increasing these variables may diminish MIA effectiveness, it concurrently enhances precise localization, thereby improving location-based services. Essentially, involving more users and increasing the number of APs benefits both localization accuracy and privacy protection. The comparison across three datasets indicates that higher numbers of users and APs lead to less successful MIA, consistent with the trends observed in Fig. 5.

5.3.4 The Impact of Classifiers in White-Box and Black-Box Settings

In this section, we explore the impact of classifier models on MIA outcomes in both white and black-box settings. Six common classification algorithms, MLP, Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and K-

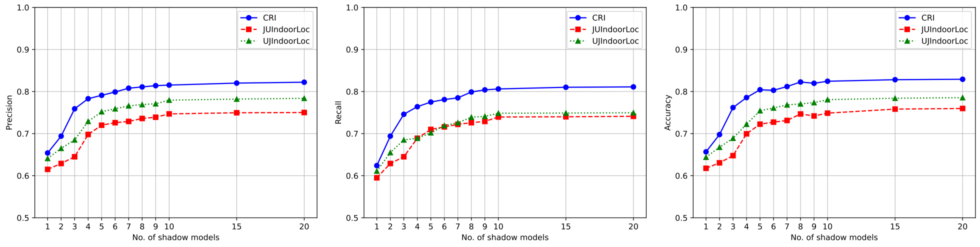
[1] https://github.com/AhmedSalem2/ML-Leaks

[2] https://github.com/csong27/membership-inference



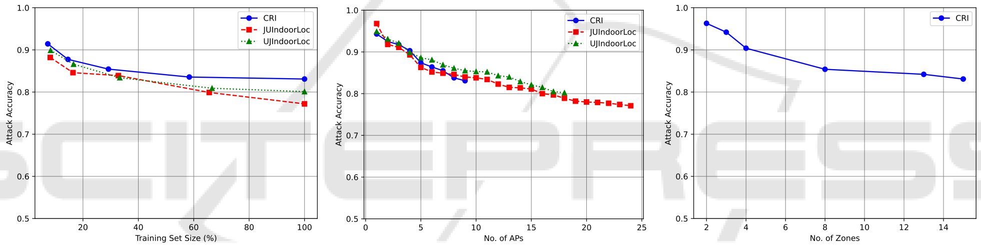
(a) Precision. (b) Recall. (c) Attack accuracy.

Figure 3: MIA performance for several numbers of epochs during training in white-box setting (10 shadow models).



(a) Precision. (b) Recall. (c) Attack accuracy.

Figure 4: MIA performance for various numbers of shadow training when in white-box setting (100 epochs).



(a) Different training sizes. (b) Different no. of APs. (c) Different no. of zones.

Figure 5: MIA accuracy when dataset size, number of features, and number of zones vary in white-box setting.

Nearest Neighbors (KNN), are used. Fig. 6 presents heatmaps illustrating MIA performance in terms of precision, recall, and accuracy for the CRI dataset. Each cell (i, j) represents MIA performance when the ith classifier serves as the target model (trained model) and the jth classifier serves as the attack model (employed by the attacker). Only diagonal cells represent MIA performance in white-box settings, where the attacker knows the model. Darker colors indicate higher MIA success. Across all classifiers, the attacker demonstrates a notable ability to identify training dataset records, with increased likelihood when utilizing MLP as either the target or attack model.

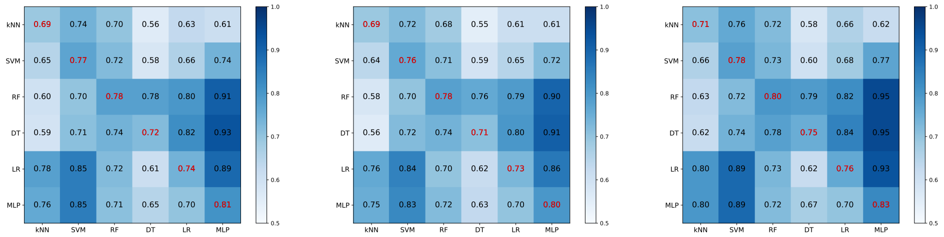
To delve deeper into classifier resilience against MIA in different settings, we conducted further experiments. In the black-box scenario, a combination of MLP, RF, and LR was used as sub-shadow models, with MLP as the target model. Conversely, the same classifier was employed in the white-box scenario. Results in Table 1 show that even in the black-box setting, the attacker achieves a success rate comparable to the white-box scenario. This highlights the vital need for robust privacy-enhancing technologies to protect individual information.

Table 1: Comparing the MIA performance in white and black box settings for CRI dataset.

Classifier	White Box			Black Box		
	Precision	Recall	Attack Accuracy	Precision	Recall	Attack Accuracy
MLP	0.81	0.80	0.83	0.83	0.79	0.84
LR	0.74	0.73	0.76	0.71	0.72	0.72
DT	0.72	0.71	0.75	0.76	0.74	0.79
RF	0.78	0.78	0.80	0.81	0.82	0.85
SVM	0.77	0.76	0.78	0.81	0.80	0.80
KNN	0.69	0.69	0.71	0.66	0.64	0.65

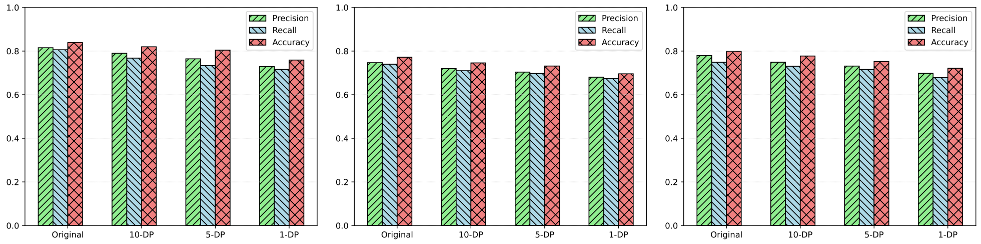
5.3.5 Defense with Differential Privacy

In this section, we use DP to defend against MIA and evaluate its effectiveness on an original indoor location dataset. We employ *PrivSyn* algorithm to generate (ε, δ)-DP datasets, with $\delta = \frac{1}{n^2}$, where n is the number of data points. We adopt $\varepsilon = 1, 5, 10$ to examine the privacy budget's impact. We aim to assess the susceptibility of the original dataset and the differentially



(a) Precision. (b) Recall. (c) Attack accuracy.

Figure 6: Heatmaps of MIA performance for the target model (rows) vs attack model (columns) for CRI dataset.



(a) CRI. (b) JUIndoorLoc. (c) UJIndoorLoc.

Figure 7: MIA performance in white-box setting before and after deploying DP.

private one to MIA and measure membership information disclosure pre- and post-DP application. This analysis shows DP's effectiveness in mitigating MIA risks and offers insights into its efficacy as a defense against privacy breaches in indoor location datasets.

Fig. 7 summarizes the white-box attack evaluation pre- and post-applying DP on original datasets. Initially, the attack achieves the highest success using original datasets for training. However, with the DP application, attack success diminishes, varying with the privacy budget. Lower budgets or increased noise lead to reduced success. Fig. 8 also compares white-box and black-box attacks before and after DP implementation on original datasets. Similar to white-box findings, the highest MIA success occurs with original data. However, applying DP notably reduces attack effectiveness, aligning with the chosen DP privacy budget. Results highlight DP's efficacy in mitigating black-box attack risks compared to white-box, strengthening the trained model privacy. These findings show that lower privacy budgets offer increased protection, emphasizing DP's importance in defending against MIA in indoor location models.

6 CONCLUSION

This study delves into protecting indoor location privacy amid the increasing use of location-based services. We assess how Membership Inference Attacks (MIA) extract sensitive indoor movement data from datasets and explore the effectiveness of Differential Privacy (DP) in mitigating these risks. Through real-world experiments, we identify vulnerabilities to MIA and analyze parameter impacts on accuracy.

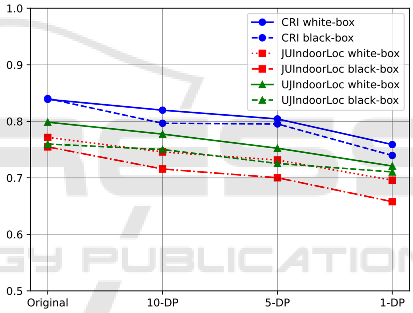


Figure 8: Comparison of MIA accuracy in white-box and black-box settings on original and DP datasets.

REFERENCES

Alhalafi, N. and Veeraraghavan, P. (2019). Privacy and security challenges and solutions in IOT: A review. *IOP Conference Series: Earth and Environmental Science*, 322:012013.

Alikhani, N., Moghtadaiee, V., Sazdar, A. M., and Ghorashi, S. A. (2018). A privacy preserving method for crowdsourcing in indoor fingerprinting localization. In *8th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 58–62.

Carlini, N., Chien, S., Nasr, M., Song, S., Terzis, A., and Tramer, F. (2022). Membership inference attacks from first principles. In *IEEE Symposium on Security and Privacy*, pages 1897–1914.

Choquette-Choo, C. A., Tramer, F., Carlini, N., and Papernot, N. (2021). Label-only membership inference attacks. In *International conference on machine learning*, pages 1964–1974.

Dionysiou, A. and Athanasopoulos, E. (2023). Sok: Membership inference is harder than previously thought. *Proceedings on Privacy Enhancing Technologies*, 3:286–306.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, page 265–284.

Fathalizadeh, A., Moghtadaiee, V., and Alishahi, M. (2023). Indoor geo-indistinguishability: Adopting differential privacy for indoor location data protection. *IEEE Transactions on Emerging Topics in Computing*, pages 1–13.

Fathalizadeh, A., Moghtadaiee, V., and Alishahi, M. (2024). Indoor location fingerprinting privacy: A comprehensive survey. *arXiv:2404.07345*.

Hayward, S., van Lopik, K., Hinde, C., and West, A. (2022). A survey of indoor location technologies, techniques and applications in industry. *Internet of Things*, 20:100608.

Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P. S., and Zhang, X. (2022). Membership inference attacks on machine learning: A survey. *ACM Computing Surveys (CSUR)*, 54(11s):1–37.

Hui, B., Yang, Y., Yuan, H., Burlina, P., Gong, N. Z., and Cao, Y. (2021). Practical blind membership inference attack via differential comparisons. *arXiv:2101.01341*.

Hyeong, J., Kim, J., Park, N., and Jajodia, S. (2022). An empirical study on the membership inference attack against tabular data synthesis models. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, page 4064–4068.

Liu, Y., Zhao, Z., Backes, M., and Zhang, Y. (2022). Membership inference attacks by exploiting loss trajectory. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 2085–2098.

Moghtadaiee, V., Ghorashi, S. A., and Ghavami, M. (2019). New reconstructed database for cost reduction in indoor fingerprinting localization. *IEEE Access*, 7:104462–104477.

Navidan, H., Moghtadaiee, V., Nazaran, N., and Alishahi, M. (2022). Hide me behind the noise: Local differential privacy for indoor location privacy. In *IEEE European Symposium on Security and Privacy Workshops*, pages 514–523.

Rahimian, S., Orekondy, T., and Fritz, M. (2020). Sampling attacks: Amplification of membership inference attacks by repeated queries. *arXiv preprint arXiv:2009.00395*.

Roy, P. and Chowdhury, C. (2021). A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(3):1–34.

Roy, P., Chowdhury, C., Ghosh, D., and Bandyopadhyay, S. (2019). Juindoorloc: A ubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity. *Wirel. Pers. Commun.*, 106(2):739–762.

Salem, A., Zhang, Y., Humbert, M., Fritz, M., and Backes, M. (2018). Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models. *ArXiv*, abs/1806.01246.

Sartayeva, Y. and Chan, H. C. (2023). A survey on indoor positioning security and privacy. *Computers & Security*, 131:103293.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy*, pages 3–18.

Song, B., Deng, M., Pokhrel, S. R., Lan, Q., Doss, R. R. M., and Li, G. (2023). Digital privacy under attack: Challenges and enablers. *ArXiv*, abs/2302.09258.

Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J. P., Arnau, T. J., Benedito-Bordonau, M., and Huerta, J. (2014). Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. In *Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 261–270.

Yang, R., Ma, J., Miao, Y., and Ma, X. (2023). Privacy-preserving generative framework for images against membership inference attacks. *IET Communications*, 17(1):45–62.

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.

Zhang, Z., Wang, T., Li, N., Honorio, J., Backes, M., He, S., Chen, J., and Zhang, Y. (2021). PrivSyn: Differentially private data synthesis. In *USENIX Security Symposium*, pages 929–946.