Co-location and Air Pollution Exposure: Case Studies on the Usefulness of Location Privacy

Liyue Fan
liyue.fan@uncc.edu
Department of Computer Science
UNC Charlotte
Charlotte, North Carolina, USA

Julius Marinak, Ashley Bang
jmarinak@uncc.edu
sbang6@uncc.edu
UNC Charlotte
Charlotte, North Carolina, USA

ABSTRACT

Location privacy has long been studied, in order to protect users' location data from untrusted servers. While existing research analyzed location privacy methods with generic utility measures, the lack of application-oriented perspectives imposes challenges for adopting location privacy. This study fills the gap by putting application utility front and center, studying the impacts of location privacy in two concrete case studies. We conduct empirical evaluations with real-world datasets from two large cities, and provide in-depth analysis on the obtained results. Furthermore, we examine the relationship between generic utility and application utility as well as the trade-off between privacy and utility in specific application settings. Our results point out interesting behaviors of the studied privacy methods and can help applications with location privacy decision-making.

CCS CONCEPTS

• **Security and privacy** → **Data anonymization and sanitization**; • **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods*; • **Information systems** → *Location based services.*

KEYWORDS

Local Privacy, Utility, Human Mobility Case Studies

ACM Reference Format:
Liyue Fan and Julius Marinak, Ashley Bang. 2022. Co-location and Air Pollution Exposure: Case Studies on the Usefulness of Location Privacy. In *The 6th ACM SIGSPATIAL Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising (LocalRec '22) (LocalRec '22), November 1, 2022, Seattle, WA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3557992.3565991

1 INTRODUCTION

In the last decades, a number of location privacy methods have been developed to hide a user's real location from untrusted servers. For instance, a recent survey [17] reviewed more than 60 location privacy methods. Our recent work [5, 6] open-sourced location privacy methods and conducted a comparative analysis using real-world trajectory datasets, with the goal of facilitating the adoption of location privacy. However, challenges remain for applications to adopt location privacy methods, especially in understanding their impacts on utility.

Existing location privacy methods are often evaluated using generic data quality measures, such as the distance between input and output locations. Unfortunately, such distances may not directly reflect the utility loss in specific applications. For example, our prior work [5] showed that a small distortion in each location record may lead to large errors in trace-level mobility patterns. Furthermore, generic measures do not take into account of the input context. For example, the same amount of distance distortion may incur different amounts of utility loss, in different spatial contexts (e.g., cities).

In this paper, we study the application utility of location privacy via two concrete case studies, which not only represent but also enable a wide range of human mobility analyses. Both studies utilize users' GPS trajectories, one detecting user pairs that co-locate with each other and the other measuring the exposure to air pollution as a user moves about the city. We conduct evaluations with real-world trajectory data and air pollution data in two large cities, to understand the practical impacts of location privacy. In addition, using the case studies, we present a correlation analysis between generic utility and application utility for the studied location privacy methods. Last but not least, we showcase how to choose location privacy methods for a specific application, by conducting a trade-off analysis between privacy and utility.

Related Work. Recently, the authors of [17] surveyed existing location privacy methods and categorized the existing methods, e.g., by architecture (i.e., trusted third party, non-trusted third party, peer-to-peer, and local) and use case (i.e., online and offline). However, the survey study on location privacy does not conduct any empirical or quantitative analysis. Our prior work [5] empirically evaluated a set of local, online privacy methods regarding utility and privacy, with real-world trajectory datasets. We considered a range of utility metrics, including distance metrics and mobility metrics, and designed two empirical privacy risk measures, i.e., re-identification and inference attacks. This study differs from our prior work [5] by focusing on the application utility for two concrete use cases, i.e., co-location and air pollution exposure. Furthermore, this study examines the relationship between generic utility (e.g., distance metrics) and application utility, as well as the tradeoff between privacy and utility, in specific application settings.

LocalRec '22, November 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9540-3/22/11…$15.00
https://doi.org/10.1145/3557992.3565991

The rest of the paper is organized as follows: Section 2 reviews the location privacy methods considered in this paper and introduces two case studies on human mobility; Section 3 introduces our evaluation methodology and presents an in-depth discussion on the results; Section 4 concludes the paper and discusses future work directions.

2 CASE STUDIES FOR LOCATION PRIVACY

Location privacy methods are often evaluated using generic data quality measures, such as the distance between input and output locations. However, generic measures may not directly reflect the utility loss in specific applications. In this section, we describe two concrete case studies on human mobility, i.e., users' GPS trajectories. Our work investigates the application utility of location privacy regarding those case studies.

2.1 Location Privacy Methods

In this work, we consider a set of existing location privacy methods which can be deployed with the *local* architecture for *online* use. We adopt the categorization of location privacy methods as in [17] and the advantages of local and online privacy methods include their strong privacy protection and the compatibility with real-time applications. Our study considers three types of location privacy methods, i.e., generalization based (Rounding/Spatial Cloaking), perturbation based (Noise/VHC/Laplace), and dummy based (SpotME/MN). The implementation for all methods evaluated in the study is publicly available[1].

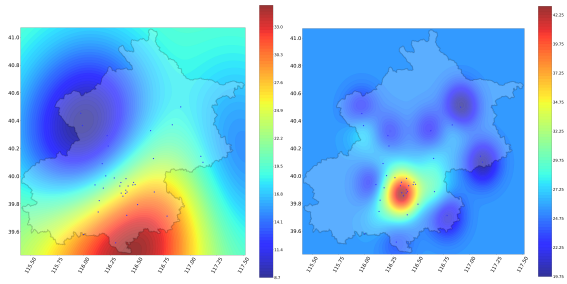
Generalization based methods hide exact location with coarse data. For example, Rounding [10, 13] snaps the input GPS coordinates to a fixed square grid with spacing s; Spatial Cloaking [10] hides a sensitive location (e.g., user's home location) inside a circular region with radius R, centered at a random location. Perturbation based methods protect user location data via modification. In this category, Noise [10] adds a random 2-D noise, the magnitude of which is drawn from a Normal distribution with variance var; VHC [16] applies space partitioning and Hilbert filling curve to the spatial domain, and perturbs the partition of the input location by drawing uniform noise from $[-\sigma, \sigma]$; Laplace [1] satisfies ϵ-geo-indistinguishability and higher ϵ indicates lower perturbation. Dummy based methods hide real locations among fake ones, i.e., dummies. SpotME [18] reports each location in the domain as the user's location with probability p; MN [9] reports multiple realistic dummy trajectories, in which adjacent locations do not deviate by more than m in latitude and longitude. Interested readers may refer to our prior work [5] for in-depth description and contrast of those methods.

2.2 Co-location

One important use case of users' location data is to detect co-locating users, i.e., individuals that are present in the same place at the same time. Users' co-locating information can benefit a wide range of applications, including digital contact tracing [21], encouraging social interaction [14], and friendship discovery [15].

In this study, we are interested in understanding the impacts of location privacy on detecting co-location. In fact, as location privacy methods may modify users' location data, the accuracy of co-location detection may be reduced. Specifically, we perform co-location detection on each user pair across all timestamps, using real data vs. sanitized data. Our empirical evaluation reports three accuracy measures: (1) false negative error - which is the total number of undetected co-locations between a pair of users; (2) false positive error - which is the total number of falsely detected co-locations between a pair of users; (3) false negative user pairs - which is the count of user pairs that have a non-zero false negative error. Note that in addition to quantifying the impacts of location privacy, those accuracy measures bear important implications to real applications. For example, a larger number of false negative user pairs may limit the efficacy of digital contact tracing during a pandemic; a high false positive error may incur high computational overheads, e.g., providing services to users who do not co-locate at the time.

[1] https://github.com/fan-group/geopriv4j



(a) $PM_{2.5}$ **(b) NO_2**

Figure 1: Air Pollution in Beijing - Interpolation from 35 Samples

2.3 Air Pollution Exposure

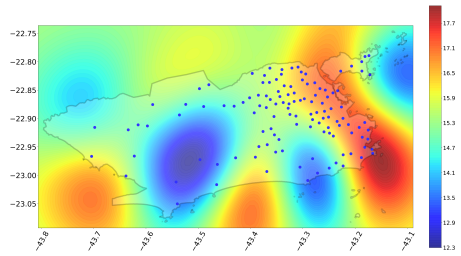
It is well demonstrated that exposure to air pollutants, such as $PM_{2.5}$ and NO_2, is associated with a range of health effects on the respiratory and other human systems [8]. As air quality varies in space and time, many research studies adopted GPS trajectories to reliably estimate the pollution levels that an individual is exposed to [2].

In our work, we aim to investigate the feasibility of adopting location privacy methods in estimating an individual's air pollution exposure, which will help address participant privacy concerns in those research studies while maintaining estimation accuracy. Specifically, we compute the absolute error in air pollutant concentrations (e.g., $PM_{2.5}$ and NO_2) using the user's real location data vs. sanitized location data, and report the average error (i.e., mean absolute error) across all timestamps. Beyond air pollution exposure, our results may shed light on other research studies which also rely on participant GPS trajectories. For example, researchers may utilize participant real-time GPS data to track an individual's exposure to neighborhood disorder and poverty, in order to predict drug cravings [4].

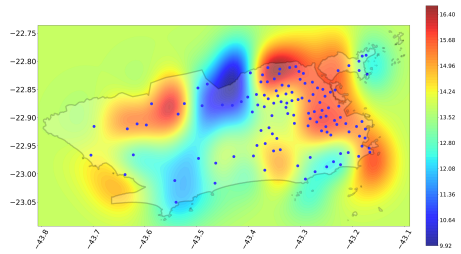
3 RESULTS AND DISCUSSIONS

3.1 Experiment Settings

Trajectories We adopt two real-world trajectory datasets, namely GeoLife [22] and RioBuses [3], and summarize the characteristics



(a) $PM_{2.5}$



(b) NO_2

Figure 2: Air Pollution in Rio de Janeiro - Interpolation from 124 Samples

of two datasets in Table 1. The GeoLife dataset records GPS data of 182 users in Beijing between 2007 and 2012. The trajectories were recorded at a variety of sampling rates, and the majority of data was logged every 1 to 5 seconds. The RioBuses dataset includes trajectories of 14149 buses in Rio de Janeiro from October 2010. The real-time GPS data was updated every minute. We select 200 buses (also referred to as users) randomly for the evaluate. Note that in both datasets, users may contribute different numbers of trajectories recorded on different dates.

Air Pollution Data. For the case study on air pollution, we are interested in understanding the effects of location privacy as individuals travel between different neighborhoods. Therefore, we consider fine grained station and neighborhood-level air quality datasets for Beijing and Rio de Janeiro, in order to incorporate spatial variability of air quality. Specifically, we adopt the observed air pollutant concentration at 35 stations in Beijing in 2014 [19], and the estimated annual average exposure data for 124 neighborhoods in Rio de Janeiro for 2001-2010 [20] and 1996-2012 [7]. Note that temporal variability of air quality is not incorporated in our experiments due to unavailability of data. In these datasets, $PM_{2.5}$ concentration is reported in micrograms per cubic meter ($\mu g/m^3$) and NO_2 is in parts per billion (ppb).

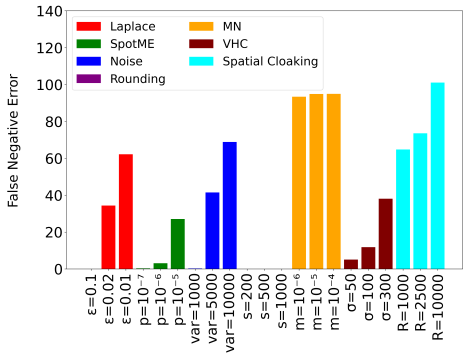
Pre-processing. For trajectory datasets, we discretize the map range of each dataset and subsample raw trajectories at 5-minute intervals. We set the user's *home* as the most frequent location at 02:00, 06:00 and 20:30 in GeoLife and as the Central bus station in RioBuses. For air pollution data, we perform kriging interpolation with the Gaussian semi-variogram to estimate the air pollutant concentrations continuously in the 2-D space from discrete samples, i.e., observations at stations. Figures 1 and 2 illustrate the interpolation results given the samples (i.e., dots in each figure) of each dataset.

Table 1: Dataset Characteristics

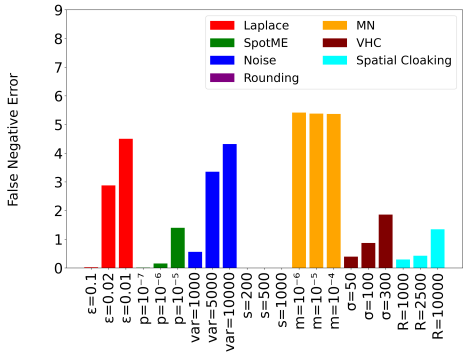
Dataset	#Users	Frequency	Resolution	Avg. # Traj's	Avg. # Loc's
GeoLife[22]	182	1 to 5 seconds	182×182	54	15640
RioBuses[3]	14149	every minute	170×170	9	2661

Table 2: Default Parameter Values

Privacy Method	Parameter	Privacy Method	Parameter
Laplace	$\epsilon = 0.02$	MN	$m = 10^{-5}$
SpotME	$p = 10^{-5}$	VHC	$\sigma = 50$
Rounding	$s = 200$	Spatial Cloaking	$R = 1000$
Noise	$var = 5000$		



(a) *GeoLife*



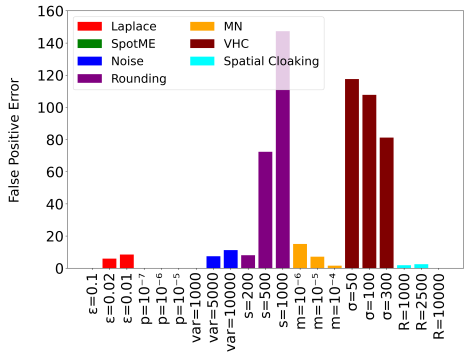
(b) *RioBuses*

Figure 3: Average False Negative Co-location Error per User Pair

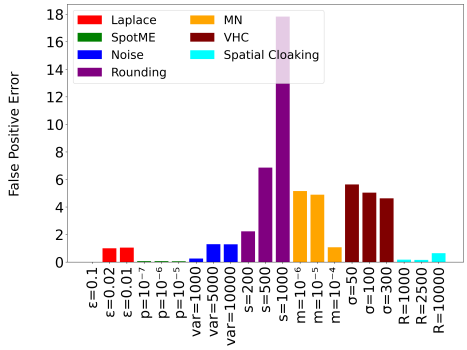
Privacy Settings All location privacy methods are implemented in Java [6]. The default parameter values are reported in Table 2.

3.2 Co-location Results

To study co-location, we report the false negative and false positive errors per user pair, averaged among all user pairs, in Figures 3 and 4. In Figure 3, we observe that all but one location privacy methods yield undetected co-locations between user pairs. The magnitude of the false negative error increases as we increase the privacy level for each privacy method. The Rounding method does not lead to false negatives, as it always reports truthful but coarser data. The MN method leads to high false negative errors in both datasets and the parameter m does not lead to significant changes, which illustrates the utility loss incurred by dummy trajectories. When comparing the two datasets, larger errors are observed in GeoLife (Figure 3a), especially for Spatial Cloaking. Intuitively, human users

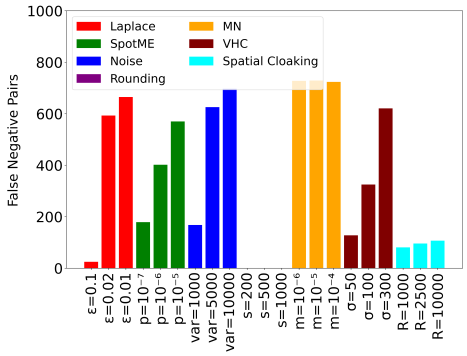


(a) *GeoLife*

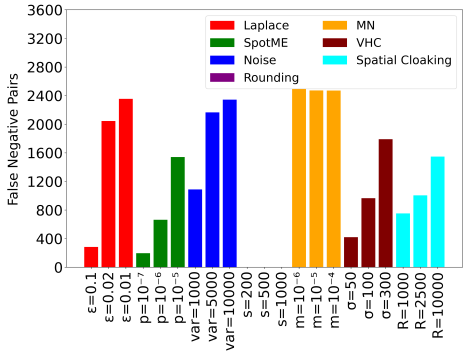


(b) *RioBuses*

Figure 4: Average False Positive Co-location Error per User Pair



(a) *GeoLife*



(b) *RioBuses*

Figure 5: False Negative User Pairs for Co-location

are mostly active near home, hence users from the same residential area often co-locate. In contrast, bus users are largely active away from home (set as the central station in Rio de Janeiro), and a pair of buses tend to have fewer co-locations as they move frequently. As a result, location privacy methods have a higher impact on human trajectories, and Spatial Cloaking, which deletes data near the user's home, inflicts larger false negative errors than other privacy methods in GeoLife.

In Figure 4, we observe low false positive errors, i.e., false co-locations, for most location privacy methods. For Laplace, SpotME, and Noise, the random perturbation mechanisms are unlikely to introduce a large number of false co-locations due to the sparsity of the geospatial data. Small false positive errors are observed for Spatial Cloaking due to the common practice of imputation, which uses the last known location to impute missing/deleted records. For Rounding, a large amount of false positives are observed to achieve a high privacy level, i.e., $s = 1000$, illustrating the trade-off between privacy and computational overheads. On the other hand, increasing the privacy level for VHC reduces the magnitude of false positive errors, as VHC utilizes a space partitioning structure to group user locations and the random perturbation (with σ param) eliminates some false positive co-locations caused by grouping.

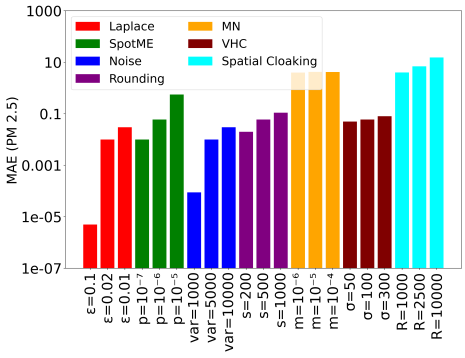
In addition to studying the average co-location error per user pair, it is crucial to understand the number of user pairs which have undetected co-locations, disregarding how many co-locations are undetected. We refer to such user pairs as "false negative pairs" and report the results in Figure 5. As expected, Rounding does not lead to any false negative user pairs. Interestingly, other privacy methods have more comparable results to each other and Spatial Cloaking yields much lower numbers of false negative user pairs in GeoLife. Contrasting Figure 3 and Figure 5, the results confirm that a small number of GeoLife users have large numbers of co-locations and a large number of RioBuses users have small numbers of co-locations.

3.3 Air Pollution Exposure Results

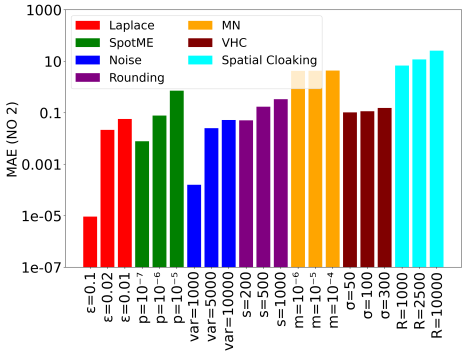
In the case of air pollution exposure, we report the MAE (mean absolute error) in exposure per time interval, averaged among all users, in Figures 6 and 7. For each method, as we increase the privacy level, larger errors are resulted. The MN method inflicts high MAE errors despite the change in the privacy parameter, similar to the generic utility results in our prior work [5], due to reporting randomly generated dummy trajectories. We observe that Laplace (with $\epsilon = 0.1$) and Noise (with $var = 1000$) lead to the lowest errors in both air pollutants and both datasets. Spatial Cloaking (with R=10000) leads to the highest errors in all cases, showcasing high impacts of data deletion in this case study. For the GeoLife dataset (Figure 6), the errors for NO_2 exposure are higher than those of $PM_{2.5}$ exposure, due to larger variations in NO_2 concentration in the dense city center (see Figure 1b).

3.4 Generic Utility vs. Application Utility

Our study investigates whether generic utility measures of location privacy can be good indicators for application utility. As a proof of

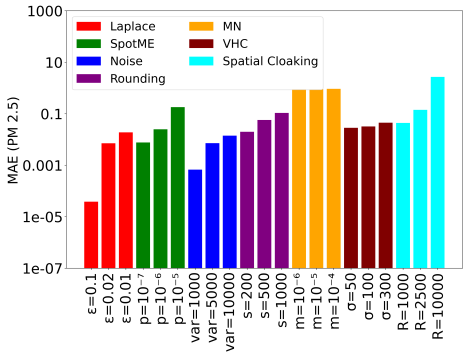


(a) $PM_{2.5}$

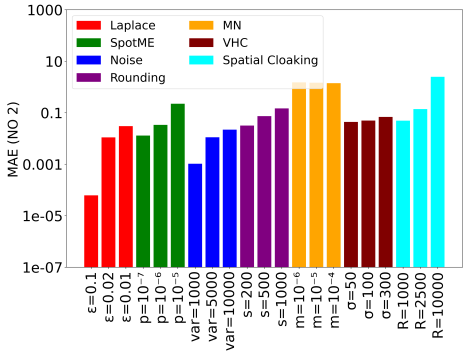


(b) NO_2

Figure 6: Errors in Air Pollution Exposure - GeoLife



(a) $PM_{2.5}$



(b) NO_2

Figure 7: Errors in Air Pollution Exposure - RioBuses

Table 3: Pearson Correlation between Air Pollution Exposure Errors and Distance-based Errors

	GeoLife		RioBuses	
	MAE NO_2	MAE $PM_{2.5}$	MAE NO_2	MAE $PM_{2.5}$
Haversine distance	0.968	0.995	0.874	0.702
Hamming distance	0.289	0.329	0.193	0.084

concept, we study the correlation between air pollution exposure errors and generic utility loss, i.e., Haversine/Hamming distances, obtained with the privacy methods under the same parameter values[2]. In Table 3, we observe that the air pollution exposure error is positively correlated with the Haversine distance, indicating that Haversine is a better utility indicator for air pollution exposure than the Hamming distance. Furthermore, the strength of correlation varies from one dataset to another, indicating that the spatial context for the input is very important, e.g., the same Haversine distance may lead to different exposure errors in different cities.

3.5 Privacy and Utility

It is important to understand that for a given application, which privacy method(s) would provide the best utility and privacy protection. To that end, we conduct a trade-off analysis as depicted in Figure 8, where each method-parameter setting is plotted based on its empirical privacy protection (x-axis) and application utility (y-axis). For brevity, we consider the utility in reporting users' $PM_{2.5}$ exposure for the GeoLife dataset, as reported in Figure 6a. To contextualize the privacy protection provided by each method-parameter setting, we consider the percentage of non re-identifiable users in the sanitized dataset, as reported in our prior work [5] ("NRI" in Figures 8 and 10). This measure indicates the level of privacy protection against a strong adversary, which may have prior knowledge about all locations visited by the target user. Higher percentages indicate stronger privacy protection, and vice versa.

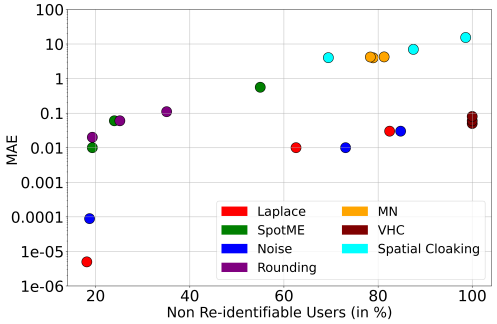


Figure 8: Trade-off between Privacy and Utility with GeoLife and $PM_{2.5}$ Exposure Errors

As shown in Figure 8, we identify three competitive methods which achieve low utility loss while providing strong privacy, namely, Laplace (red), Noise (blue), and VHC (maroon). Specifically, for any

[2]Note that we do not correlate generic utility with co-location errors. Generic utility measures only depend on individual-level data, whereas co-location errors also depend on dataset-level information, e.g., the co-locating behaviors of user pairs.

given privacy protection level (e.g., $x\%$ in NRI), the method that achieves the lowest MAE while providing equal or stronger privacy ($\geq x\%$) is one of the three consistently. Furthermore, we observe that VHC provides the strongest privacy protection (almost 100% in non re-identifiable users), and tuning its parameter value does not degrade its privacy protection or incur significantly higher utility loss. On the contrary, tuning the parameters of Laplace and Noise does provide a smooth trade-off between privacy and utility. For example, Laplace and Noise in low noise settings lead to the lowest utility loss, e.g., < 0.0001 in MAE.

4 CONCLUSION

We have presented two use cases of location privacy, i.e., co-location and air pollution exposure, with the goal of understanding the usefulness of location privacy in specific applications. We adopted real-world trajectory and air pollution data for two cities to conduct our empirical analysis. Our results in the co-location study showcase the trade-off between critical errors (i.e., undetected co-locations) and computational overheads (i.e., false positive errors). Our results in the air pollution exposure study highlight the relationship between generic utility and application utility: in this case, they are largely aligned but the alignment depends on the contexts. Furthermore, we conducted a tradeoff analysis between privacy protection and application utility, identifying three competitive location privacy methods for the specific application setting.

We consider the following directions for future work. Firstly, experimenting with different levels of spatial and temporal granularity in processing the GPS data may provide new results. For example, increasing the spatial grid resolution (thus leading to smaller cells) and estimating GPS location in 15-minute intervals may impose stronger requirements for co-location, hence resulting in different errors or behaviors in location privacy methods. Secondly, in order to incorporate temporal variability in air pollution exposure, it is possible to simulate air pollution concentrations with known diurnal and seasonal trends using Monte-Carlo simulation [11, 12]. Thirdly, it is interesting to investigate different air pollution exposure metrics, i.e., functions of concentration, used in environmental health literature. For example, our study adopted instantaneous exposure, i.e., the exposure at an instant in time, whereas other metrics include time-integrated and time-average exposures [2].

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their suggestions. This work has been supported in part by NSF CNS-1951430, CNS-2144684, and UNC Charlotte. The opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. ACM, New York, NY, USA, 901–914.

[2] Daniela Dias and Oxana Tchepel. 2018. Spatial and temporal dynamics in air pollution exposure assessment. *International journal of environmental research and public health* 15, 3 (2018), 558.

[3] Daniel S Dias, Luís Henrique MK Costa, and Marcelo Dias de Amorim. 2016. Capacity analysis of a city-wide V2V network. In *2016 7th International Conference on the Network of the Future (NOF)*. IEEE, 1–3.

[4] David H Epstein, Matthew Tyburski, Ian M Craig, Karran A Phillips, Michelle L Jobes, Massoud Vahabzadeh, Mustapha Mezghanni, Jia-Ling Lin, C Debra M Furr-Holden, and Kenzie L Preston. 2014. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. *Drug and alcohol dependence* 134 (2014), 22–29.

[5] Liyue Fan and Ishan Gote. 2021. A Closer Look: Evaluating Location Privacy Empirically. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 488–499.

[6] Liyue Fan and Sriram Yechan Gunja. 2020. Geopriv4j: an open source repository for practical location privacy. In *Proceedings of the Sixth International ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data*. 1–6.

[7] Jeffrey A Geddes, Randall V Martin, Brian L Boys, and Aaron van Donkelaar. 2016. Long-term trends worldwide in ambient NO2 concentrations inferred from satellite observations. *Environmental health perspectives* 124, 3 (2016), 281–289.

[8] Gerard Hoek, Ranjini M Krishnan, Rob Beelen, Annette Peters, Bart Ostro, Bert Brunekreef, and Joel D Kaufman. 2013. Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental health* 12, 1 (2013), 1–16.

[9] H. Kido, Y. Yanagisawa, and T. Satoh. 2005. An anonymous communication technique using dummies for location-based services. In *ICPS '05. Proceedings. International Conference on Pervasive Services, 2005.* 88–97.

[10] John Krumm. 2007. Inference Attacks on Location Tracks. In *Pervasive Computing*, Anthony LaMarca, Marc Langheinrich, and Khai N. Truong (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 127–143.

[11] Zirui Liu, Bo Hu, Lili Wang, Fangkun Wu, Wenkang Gao, and Yuesi Wang. 2015. Seasonal and diurnal variation in particulate matter (PM10 and PM2. 5) at an urban site of Beijing: analyses from a 9-year study. *Environmental Science and Pollution Research* 22, 1 (2015), 627–642.

[12] Albert M Lund, Ramkiran Gouripeddi, and Julio C Facelli. 2020. Stham: an agent based model for simulating human exposure across high resolution spatiotemporal domains. *Journal of exposure science & environmental epidemiology* 30, 3 (2020), 459–468.

[13] Kristopher Micinski, Philip Phelps, and Jeffrey S Foster. 2013. An empirical study of location truncation on android. In *IEEE Computer Society Security and Privacy Workshops Mobile Security Technologies (MoST '13)*.

[14] Thomas Olsson, Pradthana Jarusriboonchai, Paweł Woźniak, Susanna Paasovaara, Kaisa Väänänen, and Andrés Lucero. 2020. Technologies for enhancing collocated social interaction: review of design solutions and approaches. *Computer Supported Cooperative Work (CSCW)* 29, 1 (2020), 29–83.

[15] Huy Pham, Cyrus Shahabi, and Yan Liu. 2013. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*. 265–276.

[16] A. Pingley, W. Yu, N. Zhang, X. Fu, and W. Zhao. 2009. CAP: A Context-Aware Privacy Protection System for Location-Based Services. In *2009 29th IEEE International Conference on Distributed Computing Systems*. 49–57.

[17] Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, and Lionel Brunie. 2019. The Long Road to Computational Location Privacy: A Survey. *IEEE Communications Surveys Tutorials* 21, 3 (2019), 2772–2793. https://doi.org/10.1109/COMST.2018.2873950

[18] D. Quercia, I. Leontiadis, L. McNamara, C. Mascolo, and J. Crowcroft. 2011. SpotME If You Can: Randomized Responses for Location Obfuscation on Mobile Phones. In *2011 31st International Conference on Distributed Computing Systems*. 363–372.

[19] Robert A Rohde and Richard A Muller. 2015. Air pollution in China: mapping of concentrations and sources. *PloS one* 10, 8 (2015), e0135749.

[20] Aaron Van Donkelaar, Randall V Martin, Michael Brauer, N Christina Hsu, Ralph A Kahn, Robert C Levy, Alexei Lyapustin, Andrew M Sayer, and David M Winker. 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology* 50, 7 (2016), 3762–3772.

[21] Felix Nikolaus Wirth, Marco Johns, Thierry Meurers, and Fabian Prasser. 2020. Citizen-centered mobile health apps collecting individual-level spatial data for infectious disease management: scoping review. *JMIR mHealth and uHealth* 8, 11 (2020), e22594.

[22] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*. 791–800.