Privkit: A Toolkit of Privacy-Preserving Mechanisms for Heterogeneous Data Types

Data/Toolset paper

Mariana Cunha
CISUC, CRACS/INESC TEC, and Department of Computer Science, University of Porto
Porto, Portugal
mccunha@dei.uc.pt

Guilherme Duarte
CISUC, CRACS/INESC TEC, and Department of Computer Science, University of Porto
Porto, Portugal
guilherme.duarte@fc.up.pt

Ricardo Andrade
Department of Computer Science, University of Porto
Porto, Portugal
up201805015@fc.up.pt

Ricardo Mendes
CISUC and Department of Informatics Engineering, University of Coimbra
Coimbra, Portugal
rscmendes@dei.uc.pt

João P. Vilela
CISUC, CRACS/INESC TEC, and Department of Computer Science, University of Porto
Porto, Portugal
jvilela@fc.up.pt

ABSTRACT

With the massive data collection from different devices, spanning from mobile devices to all sorts of IoT devices, protecting the privacy of users is a fundamental concern. In order to prevent unwanted disclosures, several Privacy-Preserving Mechanisms (PPMs) have been proposed. Nevertheless, due to the lack of a standardized and universal privacy definition, configuring and evaluating PPMs is quite challenging, requiring knowledge that the average user does not have. In this paper, we propose a privacy toolkit - Privkit - to systematize this process and facilitate automated configuration of PPMs. Privkit enables the assessment of privacy-preserving mechanisms with different configurations, while allowing the quantification of the achieved privacy and utility level of various types of data. Privkit is open source and can be extended with new data types, corresponding PPMs, as well as privacy and utility assessment metrics and privacy attacks over such data. This toolkit is available through a Python Package with several state-of-the-art PPMs already implemented, and also accessible through a Web application. Privkit constitutes a unified toolkit that makes the dissemination of new privacy-preserving methods easier and also facilitates reproducibility of research results, through a repository of Jupyter Notebooks that enable reproduction of research results.

CCS CONCEPTS

• **Security and privacy** → **Privacy protections**; **Usability in security and privacy**; • **Software and its engineering** → *Software libraries and repositories.*

CODASPY '24, June 19–21, 2024, Porto, Portugal

ACM ISBN 979-8-4007-0421-5/24/06
https://doi.org/10.1145/3626232.3653284

KEYWORDS

Privacy, Privacy Tools, Privacy-Preserving Mechanisms, Data Privacy, Python

ACM Reference Format:
Mariana Cunha, Guilherme Duarte, Ricardo Andrade, Ricardo Mendes, and João P. Vilela. 2024. Privkit: A Toolkit of Privacy-Preserving Mechanisms for Heterogeneous Data Types: Data/Toolset paper. In *Proceedings of the Fourteenth ACM Conference on Data and Application Security and Privacy (CODASPY '24), June 19–21, 2024, Porto, Portugal.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3626232.3653284

1 INTRODUCTION

Privacy has been recognized as a fundamental human right since 1948 [31]. Nevertheless, the paced evolution of technology and the pervasiveness of continuous data collection have been raising serious privacy concerns that are often unconsidered. When sharing data, users expose sensitive information that may not always be used only for the initial purpose of the services. Moreover, people are not always aware of the potential privacy risks of sharing data and/or disregard them in detriment of benefiting from certain services [1]. In order to prevent unwanted disclosures, several Privacy-Preserving Mechanisms (PPMs) have been proposed [7] to protect privacy during the collection, flow, and storage of data.

However, selecting and configuring the proper PPM, as well as achieving the best trade-off between privacy and utility is quite challenging. Despite the efforts to clarify the privacy concept, defining privacy remains as a challenge due to the broadness of contexts to which privacy applies to. While this process of properly selecting and configuring a PPM is not trivial, it is crucial, since misconfigurations can lead to a noneffective privacy/utility level [17].

Derivative from the lack of a standardized definition of privacy is the difficulty of quantifying and assessing PPMs [18]. In an attempt to systematize this process, some frameworks have been proposed to provide a logical structure of the main components and concepts of privacy (e.g. in the context of location privacy [27, 28]). The steps for evaluating a privacy-preserving mechanism can be denoted

through a privacy pipeline, where the simplest pipeline can consist of the raw data, followed by a defense mechanism, and the resulting private data. From this, the achieved level of privacy and utility can be quantified. The analysis can then be complemented by applying an attack and computing relevant metrics to assess the privacy-utility trade-off. However, existing metrics are often specific to a data type application and, thus, measuring and comparing the privacy/utility level achieved by mechanisms is still challenging.

In this paper, we propose a privacy toolkit - Privkit[1] - to address the identified challenges by systematizing the process of quantifying the privacy level for various data types. From the analysis of existing frameworks/tools that allow to test and configure PPMs (e.g. *ARX* [22] and *Accio* [23]), we concluded that the majority is focused on a specific data type and, consequently, on specific PPMs. The objective of this work is not only systematizing the process of quantifying privacy, but also creating a privacy toolkit that is open source, extendable, and that supports heterogeneous data types, as well as appropriate PPMs, attacks, and metrics. This toolkit is available through a Python Package with several state-of-the-art mechanisms already implemented, and also accessible through a Web application. Privkit constitutes a unified toolkit that makes the dissemination of new algorithms easier. Furthermore, this toolkit is a contribution towards mitigating the limited availability of mechanisms' source code, which will also facilitate the reproducibility of results by providing a repository where the research community can detail the performed experiments.

The remainder of this paper is structured as follows. Section 2 provides an overview of background concepts and related work by presenting existing frameworks and tools. Section 3 presents the proposed toolkit - Privkit. Section 4 details the Privkit in practice through use cases and Section 5 draws the main conclusions.

2 BACKGROUND AND RELATED WORK

The large amounts of data that are continuously collected have resulted in serious risks to privacy. In order to address these risks, several Privacy-Preserving Mechanisms (PPMs) have been proposed with the aim of preventing unwanted disclosures. Nonetheless, the selection of PPMs should take into account the data types under consideration, their sensitivity, and the application in which they are being used, as different applications will require different PPMs.

Privacy-preserving mechanisms can be divided into different categories depending on the type of use (online/offline), but also on their approach (e.g. anonymization or obfuscation). Furthermore, the selection of a PPM should be performed in accordance to the data type, as detailed in the privacy taxonomy of [7]. This selection is, however, quite challenging, as is its configuration, due to the diversity of approaches and requiring parameters. To evaluate the chosen mechanism and in order to guarantee an adequate privacy-utility trade-off, it is important to rely on proper privacy and utility metrics, such as the correctness and quality loss, respectively [18].

Due to the lack of a standardized privacy definition, quantifying and measuring privacy is not been trivial. In particular, the fact that current landscape of privacy-preserving mechanisms require configuring parameters that can either be hard to define (e.g. the meaning of epsilon in differential privacy [14]), or recalculated for each environment (e.g. k parameter in k-anonymity [24]) imposes a difficulty during the configuration step. This is especially critical, since a misconfiguration can lead to a noneffective privacy/utility [17]. To facilitate the test of different configurations and systematize the evaluation of PPMs and/or comparison among mechanisms, some frameworks/tools have been proposed [7], as detailed next.

Selecting a PPM that is appropriate to a data type and data application is a fundamental aspect. In this sense, the existing tools tend to be focused on only one type/format of data and, consequently, on specific PPMs for that type of data, which presents a limitation. In the context of structured data, several tools have been developed, either focused on tabular data (i.e. data stored in tables), such as *ARX* [22] or *Anonimatron*[2], or microdata, such as *μ-ARGUS*, *τ-ARGUS*, and *sdcMicro* from *Statistical Disclosure Control Tools*[3]. Beyond tabular data, the *Amnesia*[4] tool also handles set-valued data. In addition, the demand for location privacy has led to the development of tools focused on mobility data, such as: *Location Privacy Meter* [26], *GEPETO* [10], and *Accio* [23]. Concerning visual data, there is an extensive literature on PPMs for 2-dimensional data types [15], but little to none tools for privacy of point cloud data, an emerging data type for storage of 3-dimensional data. However, widely used libraries like Open3D [36], Trimesh[5], and PyVista [30] support 3D data processing. With respect to the implemented PPMs, as a consequence of the supported data types, in general, tools that are focused on structured data implement anonymization mechanisms, whereas location privacy tools focus on obfuscation mechanisms. Concerning the achieved trade-off between privacy and utility, most frameworks/tools provide metrics for privacy evaluation, but some lack the utility assessment [7]. This is a crucial drawback as the selection of a mechanism should weigh this trade-off.

Building on the limitations identified in the existing frameworks/-tools, we propose a novel privacy toolkit that is suitable for different data types, privacy-preserving mechanisms, metrics, and attacks. This toolkit is easily accessible through a Python Package and allows the evaluation of both privacy and utility. Moreover, the fact that Privkit is open source and extendable makes it an incentive to the dissemination of algorithms from the research community, and also contributes to the possibility of reproducing results. The design of Privkit is detailed in the next section.

3 PRIVKIT DESIGN AND ARCHITECTURE

Privkit is a novel privacy toolkit that arises from the need of quantifying and measuring privacy in a systematized manner. In addition, the lack of a tool that standardizes the assessment of Privacy-Preserving Mechanisms (PPMs) for heterogeneous data types also motivated the development of an open source toolkit. In light of this, Privkit follows a modular programming approach, where the typical steps of a privacy pipeline are designed as modules. For illustration purposes, Figure 1 presents an example of a privacy pipeline with the following components: data processing, PPM, attack, and privacy/utility evaluation. In this way, by structuring these main components, Privkit standardizes the privacy pipeline

[1] https://privkit.fc.up.pt/

[2] https://github.com/realrolfje/anonimatron/
[3] https://github.com/sdcTools/
[4] https://amnesia.openaire.eu/
[5] https://trimesh.org/

to test, configure, and evaluate mechanisms, while facilitating the implementation of new data types, PPMs, attacks, and metrics.

Figure 1: An example of a privacy pipeline with the application of a privacy-preserving mechanism, an attack, and the evaluation of the privacy and utility level of data.

Figure 2 presents an overview of the architecture, where each module is presented with the respective abstract class (*Adapter*) that can be used to extend the toolkit with the implementation of new features. These *Adapters* define the structure of the source code in accordance to the corresponding module, as well as the methods to be implemented. This emphasizes the modularity of the toolkit, while fostering the code readability, maintainability, and scalability. The current version of Privkit has implemented methods for processing two data types, and a set of 11 PPMs, 6 attacks, and 9 metrics (see some selected examples in Figure 2). Apart from that, the toolkit also provides a module to access datasets and a module with utility methods. The four main components and available implementations are detailed next.

- **Data Adapter:** defines the structure to handle a data type by providing methods to load, process and save data. Location data and facial data are the current available data types handled by the toolkit.
- **PPM Adapter:** defines the structure of a privacy-preserving mechanism with an executor method to be implemented. The current version of Privkit provides the following PPMs in the context of location data: Geo-indistinguishability (Geo-Ind) [4], Adaptive Geo-Ind [2], Clustering Geo-Ind [6], VA-GI [16], and Privacy-Aware Remapping [9]. Within facial data, the supported PPMs are CentroidVoxel, UniformNoise, Tapering, SmoothKNN, Merge2Faces, and Point-Mesh-Point (PMP) [3].
- **Attack Adapter:** defines the structure of an adversary attack with an executor method to be implemented. The current version of the Privkit provides the following attacks in the context of location data: Map-Matching [12, 19], two Optimal Attacks [29], Profile-Estimation Based Attack (PEBA) [20], and Top-N Attack [33]. Within facial data, a 2D face recognition model is employed, offering the flexibility to choose from a variety of models within the Deepface face recognition framework [25]. Both biometric modes, verification and identification are available for testing the PPMs.
- **Metric Adapter:** defines the structure of a metric with an executor method to be implemented. The current version of the Privkit holds several metrics to measure the achieved privacy and utility levels, such as: adversary error, quality loss, and F1 Score. Within facial data, the supported privacy and utility metrics include the Cumulative Matching Characteristic (CMC) curve, the Receiver Operating Characteristic (ROC) curve and respective AUC value, the Fréchet Inception Distance (FID), the Structural Similarity Index Measure (SSIM), and the face detection accuracy [3].

To make this toolkit publicly available and facilitate its usage with a minimal learning curve, Privkit is designed as a Python Package that can be easily imported as a library and used in a development project. The fact that there are well-known scientific tools provided in Python within different contexts (e.g. machine learning) also justifies the choice for developing this open source toolkit in Python. Similarly to other Python Packages, Privkit is available in a Github repository[6] to foster contributions from the research community, while maintaining version control and code quality. The contributing guidelines are detailed in the repository, explaining the process of identifying/reporting issues and the review workflow of contributions through a pull-based model [11, 32].

On the other hand, to foster privacy in a wide manner for a general audience beyond the research community, Privkit is also accessible through a Web application. This works as a demonstrator, where the implementations provided in the Python Package are automatically provided as a Web application. In the toolkit Web version, users can select the proper mechanism, test configurations, and evaluate through existing metrics in a limited sample of the data. The configurations performed and tested in the Web version can then be downloaded as a Python script to execute locally.

With the last requirement of promoting the reproducibility of research results, Privkit constitutes a unified toolkit that provides a repository of Jupyter Notebooks[7]. This enables the exact reproduction of results from research papers. The current version of the toolkit has already available the reproducibility of experiments from existing works, such as [9, 16]. This paper itself also has a corresponding Jupyter Notebook to allow the reproduction of the demonstration results presented in the next section.

4 PRIVKIT IN PRACTICE – USE CASES

This section demonstrates Privkit in practice by relying on two use cases: location data (Section 4.1) and facial data (Section 4.2). The objective of these use cases is to exemplify the implementation of two different data types and the experiments that can be performed through the toolkit. For each use case, we detail the considered privacy pipeline, the related source code, and the results of the targeted assessments. The following sections assume that the Privkit library is installed and has already been imported, which is as simple as presented in Listing 1.

```
1 import privkit as pk
```

Listing 1: Import Privkit Package.

4.1 Use Case 1: Location Data

Location privacy is an emerging topic of research due to the pervasiveness of Location-Based Services (LBSs) in always mobile devices. Regardless of the benefits for the user, this flow of information poses a threat to users' location privacy, since location data allows to determine the users' identity, habits, social relationships or even health conditions [5, 8, 13]. To mitigate the existing risks, several Location Privacy-Preserving Mechanisms (LPPMs) have been proposed. Taking this into consideration, we now present a use case related to location data.

By extending the Data Adapter, Privkit supports a new component that implements methods to load, process, and store the

[6] https://github.com/privkit/privkit
[7] https://github.com/privkit/privkit-tutorials

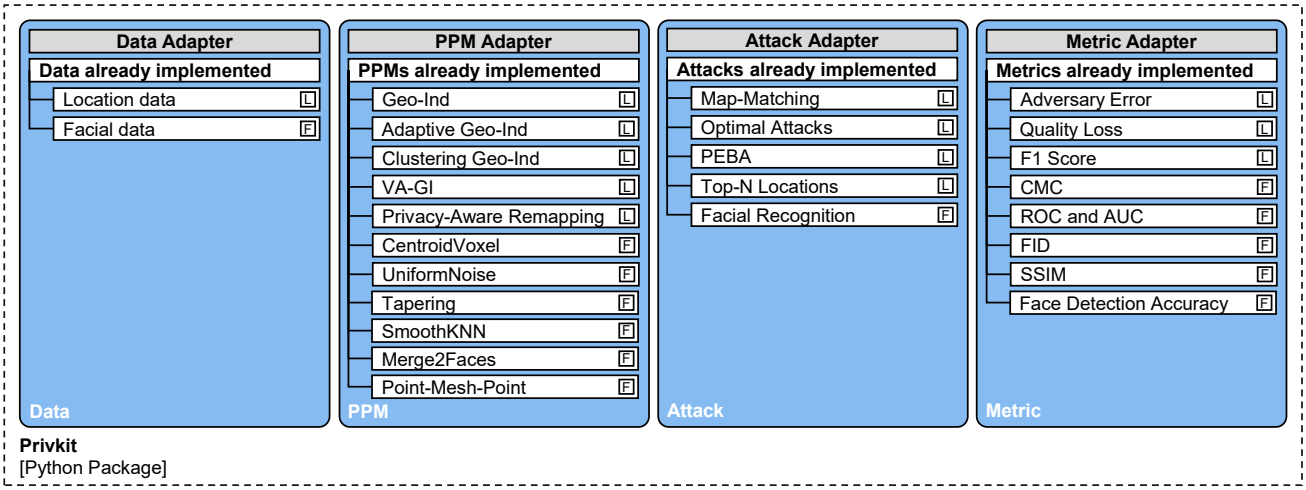


Figure 2: Privkit architecture overview with a representation of the available modules and selected examples of already implemented mechanisms.

location data type. A similar process is due for each novel data type to be implemented, as well as for each of the remaining Adapters (PPM, Attack, and Metric). A tutorial and more examples on how to add new data types to Privkit is available in the toolkit documentation. Location data consists of any pair <latitude, longitude> and optional features, such as the datetime of data collection. In this case, the data processing includes methods that can be used, for instance, to filter by coordinates within a bounding-box, but also to remap the locations into a grid. To facilitate the test of this component, as well as the test of LPPMs with real-world data, this toolkit makes available the access and pre-processing of two popular datasets: Geolife [35] and Cabspotting [21]. Therefore, the following listings of code represent examples of the first step execution from our privacy pipeline, that is, the data component. Listing 2 presents an example of loading location data from an array, nonetheless, as long as the position of the required features is indicated (i.e. latitude and longitude), location data can be loaded from other formats, such as a file (e.g. Comma-Separated Value (CSV) or Pickle) or a Pandas dataframe. Listing 3 presents an example of loading data from a dataset, in this case, the Geolife dataset. A sample of this dataset constitutes the target of analysis in this use case.

Considering the privacy pipeline, we now detail the component related to the application of a privacy-preserving mechanism. Within the context of location data, we provide several state-of-the-art mechanisms (see Section 3) and each mechanism has information about the research paper that proposed such method, where more in-depth details can be consulted. In order to provide a ready-to-use version of the implemented mechanisms, the required configuration parameters have a default value in the available implementation. In this paper, we present the application and comparison of the Geo-Ind [4] and Clustering Geo-Ind [6] mechanisms with a privacy parameter epsilon of 0.016 and radius value of 100 meters defined for illustration purposes. Each implemented mechanism extends the PPM component with its own execution method.

Listing 4 demonstrates the possibility of testing several configurations, as well as the modularity of this toolkit, which makes it easy to integrate into an existing project. The applied PPM then returns the private data and is also able to compute the quality loss metric. This metric measures the achieved level of data utility, which, in this case, corresponds to the distance between the exact user's location and the obfuscated location.

```
data_to_load = [['2008-10-23 02:53:04', 39.984702, 116.318417],
                ['2008-10-23 02:53:10', 39.984683, 116.31845],
                ['2008-10-23 02:53:15', 39.984686, 116.318417]]

location_data = pk.LocationData()
location_data.load_data(data_to_load, datetime=0, latitude=1,
↪  longitude=2)
```

Listing 2: Example of loading data from an array with datetime, latitude, and longitude.

```
geolife_dataset = pk.GeolifeDataset()
geolife_dataset.load_dataset()

location_data = geolife_dataset.data
```

Listing 3: Example of loading data from the dataset module, in this case, the Geolife dataset [35].

```
planar_laplace = pk.PlanarLaplace(epsilon=0.016)
obfuscated_data_pl = planar_laplace.execute(location_data)

clustering = pk.ClusteringGeoInd(epsilon=0.016, r=100)
obfuscated_data_cgi = clustering.execute(location_data)
```

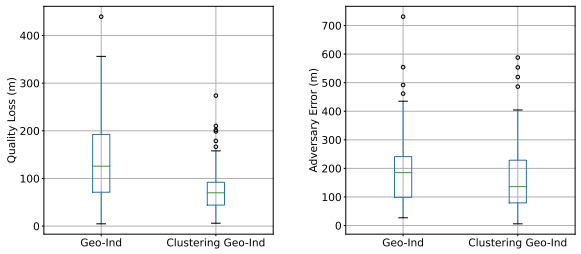
Listing 4: Example of applying Geo-Ind and Clustering Geo-Ind mechanisms with a privacy parameter epsilon of 0.016 and a radius value of 100 meters.

To demonstrate a complete privacy pipeline, we now apply an attack. The Privkit has already implemented state-of-the-art mechanisms that, similarly to PPMs, are an extension of the Attack Adapter with their own execution method to be implemented. In this demonstration, we resort to the Map-Matching (MM) attack [12, 19] to simulate an adversary and rely on the adversary error to measure the achieved privacy level. The adversary error measures the distance between the exact user's location and the adversary estimation of the user's location. Figure 3 presents the privacy/utility results from the executed privacy pipeline, where we applied the

Geo-Ind and Clustering Geo-Ind PPMs and the MM attack. From the results, we can assess the PPMs in terms of utility, with the quality loss metric, and privacy, with the adversary error metric, which allows us to observe a better privacy-utility trade-off from the Clustering Geo-Ind in comparison to Geo-Ind, since it maintains the achieved privacy level of data, while decreasing the quality loss.

```
map_matching = pk.MapMatching(G=road_network)
adversary_data = map_matching.execute(location_data)

adversary_error = pk.AdversaryError()
adversary_error.execute(adversary_data)
```

Listing 5: Example of applying the Map-Matching attack with a given road network and computing the adversary error.



(a) Quality Loss. **(b) Adversary Error.**

Figure 3: Boxplot of the quality loss, in meters, for the Geo-Ind and Clustering Geo-Ind PPMs, and the adversary error, in meters, measured after applying the Map-Matching attack.

4.2 Use Case 2: Facial Data

Facial privacy remains a focal point of ongoing research, fueled by the rapid expansion of facial recognition technology and the widespread collection of facial data [34]. While strides have been made in addressing 2D privacy-preserving solutions, a clear absence exists in their development for the 3D space. To bridge this gap, Privkit currently offers support for 3D facial data, and we present a use case to underscore its practical application.

Building upon the Data Adapter, Privkit manages 3D facial data represented as point clouds. This data format consists of a discrete set of points defined by <x, y, z> coordinates, with the ability to incorporate color information for individual data points. Mobilizing the data component, we can load, process, and store facial data. Listing 6 illustrates loading facial data from a file where the standard file types such as Polygon File Format (PLY) and Point Cloud Data (PCD) are supported. Alternatively, data can be loaded from an array. The data processing encompasses a range of methods for segmenting facial points using a bounding box or a sphere with a predefined center and radius, eliminating outliers, and downsampling a point cloud. The toolkit also provides access to a sample of a custom-made facial dataset with ten identities that includes both 3D and 2D information, aiding in the testing of the data component.

In relation to the upcoming stage in the privacy pipeline, which involves the implementation of privacy-preserving mechanisms, we provide six mechanisms with distinct principles (see Section 3). Each mechanism includes information from the relevant research paper, offering insights into its functionality and configuration setup. In this paper, we demonstrate the application of one mechanism, the Point-Mesh-Point, using a predefined set of regulating parameters for illustration purposes as the application of the remaining is analogous (refer to Listing 7). Each implemented mechanism extends the PPM component and comes with its own execution method.

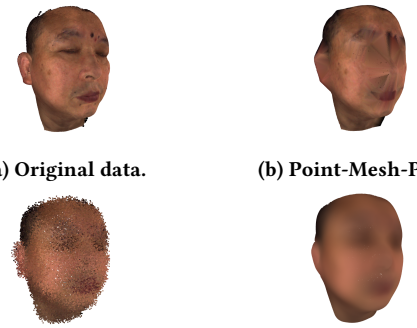
```
data_to_load = 'sample_face.ply'

facial_data = pk.FacialData()
facial_data.load_data(data_to_load)
```

Listing 6: Example of loading data from a PLY file.

```
point_mesh_point = pk.PointMeshPoint(alpha=40, n=200000)
obfuscated_data_PMP = point_mesh_point.execute(facial_data)
```

Listing 7: Example of applying the Point-Mesh-Point mechanism with a set of illustrative parameters.

To evaluate the achieved level of privacy, we can compute relevant metrics. For example, we have metrics implemented for measuring the attained privacy level, such as the CMC curve that measures the cumulative probability of correctly identifying a subject across different rank positions (identification mode), and the ROC curve and its AUC that quantify the trade-off between true positive rate and false positive rate (verification mode). As for data utility, three metrics associated with human face resemblance, image quality, and human perception are employed. Respectively, these metrics are determined by the accuracy of a face detector in identifying a human face on the anonymized data, the SSIM, and the FID. Both privacy and utility evaluation procedures are conducted within the 2D space using an orthographic projection of the 3D point clouds, taking advantage of the more mature development in the biometry field. However, other than the providing results for these metrics, for the purpose of this facial data type, we provide the output of the anonymization process, as illustrated in Figure 4.



(a) Original data. **(b) Point-Mesh-Point.**

(c) UniformNoise. **(d) SmoothKNN.**

Figure 4: Visual comparison of three facial PPMs with selected example configurations[8].

5 CONCLUSION

The ubiquity of smart devices has contributed to collect large amounts of data. Despite the possible benefits arising from this data analysis, the potential privacy risks are also evident. Therefore, implementing Privacy-Preserving Mechanisms (PPMs) is imperative to prevent privacy breaches. Nonetheless, selecting and configuring

[8]The 3D model "face raw Mesh" by Diana Liu is available under Creative Commons Attribution at https://skfb.ly/6Ut7w

the appropriate PPM can become a challenging task due to the lack of a systematized manner for defining and evaluating privacy. This work responds to this problem by proposing Privkit, a privacy toolkit to test PPMs and quantify the achieved privacy/utility level. This tool is open source and available both as a Python Package and a Web application. As future work, an API will also be available. The current version of the toolkit provides state-of-the-art mechanisms in the context of location data and facial data, and aims to be extended to any type of data and implement new privacy-preserving mechanisms, attacks, and metrics. Privkit is an incentive for the research community to make their algorithm proposals publicly available and, thus, contribute to the results reproducibility.

ACKNOWLEDGMENTS

This work is funded by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020. This work was performed in the scope of the Project Theia [Project nº 047264, Funding Reference POCI-01-0247-FEDER-047264], funded by European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) and Portugal 2020, and in the scope of the Smart Networks and Services Joint Undertaking (SNS JU) under the EU Horizon Europe programme PRIVATEER under Grant Agreement No. 101096110. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or SNS JU. Mariana Cunha wishes to acknowledge financial support by the Portuguese funding institution Fundação para a Ciência e a Tecnologia (FCT) under the grant 2020.04714.BD.

REFERENCES

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[2] Raed Al-Dhubhani and Jonathan M Cazalas. 2017. An adaptive geo-indistinguishability mechanism for continuous LBS queries. *Wireless Networks* 24 (2017), 3221–3239.

[3] Ricardo Andrade. 2023. *Privacy-Preserving Face Detection: A Comprehensive Analysis of Face Anonymization Techniques.* Master's thesis. University of Porto.

[4] 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* (Berlin, Germany) *(CCS '13)*. Association for Computing Machinery, New York, NY, USA, 901–914.

[5] Benjamin Baron and Mirco Musolesi. 2020. Where you go matters: a study on the privacy implications of continuous location tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–32.

[6] Mariana Cunha, Ricardo Mendes, and João P. Vilela. 2019. Clustering Geo-Indistinguishability for Privacy of Continuous Location Traces. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)*. IEEE, 1–8. https://doi.org/10.1109/CCCS.2019.8888111

[7] Mariana Cunha, Ricardo Mendes, and João P Vilela. 2021. A survey of privacy-preserving mechanisms for heterogeneous data types. *Computer Science Review* 41 (2021), 100403.

[8] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3, 1 (2013), 1–5.

[9] Guilherme Duarte, Mariana Cunha, and João P. Vilela. 2024. A Privacy-Aware Remapping Mechanism for Location Data. In *The 39th ACM/SIGAPP Symposium on Applied Computing* (Avila, Spain) *(SAC '24)*. Association for Computing Machinery, New York, NY, USA.

[10] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Nuñez Cortez. 2010. Gepeto: a geoprivacy-enhancing toolkit. In *2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops*. IEEE, 1071–1076.

[11] Georgios Gousios, Martin Pinzger, and Arie van Deursen. 2014. An Exploratory Study of the Pull-Based Software Development Model. In *Proceedings of the 36th International Conference on Software Engineering* (Hyderabad, India) *(ICSE 2014)*. Association for Computing Machinery, New York, NY, USA, 345–355.

[12] George R Jagadeesh and Thambipillai Srikanthan. 2017. Online map-matching of noisy and sparse location data with hidden Markov and route choice models. *IEEE Transactions on Intelligent Transportation Systems* 18, 9 (2017), 2423–2434.

[13] John Krumm. 2009. A survey of computational location privacy. *Personal and Ubiquitous Computing* 13, 6 (2009), 391–399.

[14] Jaewoo Lee and Chris Clifton. 2011. How much is enough? choosing ε for differential privacy. In *International Conference on Information Security*. Springer, 325–340.

[15] Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J. Scheirer, Arun Ross, Peter Peer, and Vitomir Štruc. 2021. Privacy–Enhancing Face Biometrics: A Comprehensive Survey. *IEEE Transactions on Information Forensics and Security* 16 (2021), 4147–4183. https://doi.org/10.1109/TIFS.2021.3096024

[16] Ricardo Mendes, Mariana Cunha, and João P. Vilela. 2023. Velocity-Aware Geo-Indistinguishability. In *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy* (Charlotte, NC, USA) *(CODASPY '23)*. Association for Computing Machinery, New York, NY, USA, 141–152.

[17] Ricardo Mendes, Mariana Cunha, and João P Vilela. 2020. Impact of frequency of location reports on the privacy level of geo-indistinguishability. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (2020), 379–396.

[18] Ricardo Mendes and João P. Vilela. 2017. Privacy-Preserving Data Mining: Methods, Metrics, and Applications. *IEEE Access* 5 (June 2017), 10562–10582.

[19] Paul Newson and John Krumm. 2009. Hidden Markov Map Matching through Noise and Sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Seattle, Washington) *(GIS '09)*. Association for Computing Machinery, New York, NY, USA, 336–343.

[20] Simon Oya, Carmela Troncoso, and Fernando Pérez-González. 2019. Rethinking location privacy for unknown mobility behaviors. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 416–431.

[21] Michal Piorkowski, Natasa Sarafijanovic-Djukic, and Matthias Grossglauser. 2022. CRAWDAD epfl/mobility. https://doi.org/10.15783/C7J010 Accessed: Dec 2023.

[22] Fabian Prasser, Florian Kohlmayer, Ronald Lautenschläger, and Klaus A Kuhn. 2014. Arx-a comprehensive tool for anonymizing biomedical data. In *AMIA Annual Symposium Proceedings*, Vol. 2014. American Medical Informatics Association, 984.

[23] Vincent Primault, Mohamed Maouche, Antoine Boutet, Sonia Ben Mokhtar, Sara Bouchenak, and Lionel Brunie. 2018. ACCIO: How to Make Location Privacy Experimentation Open and Easy. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 896–906.

[24] Pierangela Samarati and Latanya Sweeney. 1998. *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.* Technical Report. technical report, SRI International.

[25] Sefik Ilkin Serengil and Alper Ozpinar. 2020. LightFace: A Hybrid Deep Face Recognition Framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 23–27.

[26] Reza Shokri and Vincent Bindschaedler. 2021. Location Privacy Meter. https://github.com/privacytrustlab/location_privacy_meter. Accessed: Dec 2023.

[27] Reza Shokri, Julien Freudiger, and Jean-Pierre Hubaux. 2010. A Unified Framework for Location Privacy. http://infoscience.epfl.ch/record/148708

[28] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. 2011. Quantifying location privacy. In *2011 IEEE symposium on security and privacy*. IEEE, 247–262.

[29] Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. 2012. Protecting Location Privacy: Optimal Strategy against Localization Attacks. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security* (Raleigh, North Carolina, USA) *(CCS '12)*. Association for Computing Machinery, New York, NY, USA, 617–627.

[30] Bane Sullivan and Alexander Kaszynski. 2019. PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *Journal of Open Source Software* 4, 37 (May 2019), 1450.

[31] The United Nations. 1948. Universal Declaration of Human Rights.

[32] Yue Yu, Huaimin Wang, Gang Yin, and Tao Wang. 2016. Reviewer recommendation for pull-requests in GitHub: What can we learn from code review and bug assignment? *Information and Software Technology* 74 (2016), 204–218.

[33] Hui Zang and Jean Bolot. 2011. Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MobiCom '11)*. Association for Computing Machinery, New York, NY, USA, 145–156.

[34] Yushu Zhang, Tao Wang, Ruoyu Zhao, Wenying Wen, and Youwen Zhu. 2023. RAPP: Reversible Privacy Preservation for Various Face Attributes. *IEEE Transactions on Information Forensics and Security* 18 (2023), 3074–3087.

[35] Yu Zheng, Hao Fu, Xing Xie, Wei-Ying Ma, and Quannan Li. 2011. Geolife GPS trajectory dataset - User Guide. https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/ Accessed: Dec 2023.

[36] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. 2018. Open3D: A Modern Library for 3D Data Processing. *arXiv:1801.09847* (2018).